



经验证的 NetApp 架构

## 由 NVIDIA 提供支持的 NetApp ONTAP AI

可扩展的人工智能基础架构：专为实际深度学习使用情形而设计

NetApp 公司的 David Arnette、Sundar Ranganathan、Amit Borulkar、Sung-Han Lin 和 Santosh Rao

合作方



## 目录

<b>1</b>	<b>内容提要 .....</b>	<b>1</b>
<b>2</b>	<b>计划摘要 .....</b>	<b>1</b>
2.1	经验证的 NetApp 架构计划 .....	1
2.2	NetApp ONTAP AI 解决方案 .....	1
<b>3</b>	<b>深度学习数据管道 .....</b>	<b>2</b>
<b>4</b>	<b>解决方案概述 .....</b>	<b>3</b>
4.1	解决方案技术 .....	3
4.2	NVIDIA DGX-1 服务器 .....	4
4.3	NetApp AFF 系统 .....	4
4.4	NetApp ONTAP 9 .....	5
4.5	NetApp FlexGroup 卷 .....	5
4.6	NVIDIA GPU Cloud 和 Trident .....	6
4.7	Cisco Nexus 3232C 网络交换机 .....	6
4.8	基于融合以太网的 RDMA .....	7
<b>5</b>	<b>技术要求 .....</b>	<b>7</b>
5.1	硬件要求 .....	7
5.2	软件要求 .....	7
<b>6</b>	<b>解决方案架构 .....</b>	<b>8</b>
6.1	网络拓扑结构和交换机配置 .....	8
6.2	存储系统配置 .....	9
6.3	主机配置 .....	10
<b>7</b>	<b>解决方案验证 .....</b>	<b>11</b>
7.1	验证测试计划 .....	11
7.2	验证测试结果 .....	12
7.3	解决方案规模估算指导 .....	16
<b>8</b>	<b>结论 .....</b>	<b>16</b>
	<b>致谢 .....</b>	<b>16</b>
	<b>如何查找其他信息 .....</b>	<b>16</b>
	<b>附录 .....</b>	<b>iv</b>
	每个模型使用不同批大小所达到的训练速度 .....	iv
	每个模型的 GPU 扩展比较 .....	iv
	Tensor 核心和 CUDA 核心的比较 .....	v
	所有模型的 GPU 工作负载 .....	v

## 表格目录

表 1) 硬件要求。 .....	7
表 2) 软件要求。 .....	7

## 插图目录

图 1) NetApp ONTAP AI 解决方案机架级架构。 .....	2
图 2) 从边缘到核心再到云端的数据管道。 .....	2
图 3) 经验证的 NetApp ONTAP AI 解决方案架构。 .....	4
图 4) NetApp FlexGroup 卷。 .....	6
图 5) 为融合增强型以太网标准以及 RoCE v1 和 v2 提供 NX-OS 支持的 Cisco Nexus 交换机。 .....	6
图 6) 网络交换机端口配置。 .....	8
图 7) DGX-1 和存储系统端口的 VLAN 连接。 .....	9
图 8) 存储系统配置。 .....	10
图 9) DGX-1 主机的网络端口和 VLAN 配置。 .....	11
图 10) 所有模型的训练吞吐量。 .....	13
图 11) GPU 利用率和存储带宽 (VGG16)。 .....	13
图 12) 所有模型的推理 (Tensor 核心和 CUDA 核心)。 .....	14
图 13) 所有模型的存储带宽。 .....	15
图 14) 所有模型的存储延迟。 .....	15
图 15) 所有模型的存储 CPU 利用率。 .....	15
图 16) 训练模型使用各种不同批大小时的比较。 .....	iv
图 17) 各种训练模型的 GPU 扩展。 .....	v
图 18) CUDA 核心与 Tensor 核心之间的性能比较。 .....	v
图 19) ResNet-50 的 GPU 利用率和存储带宽。 .....	vi
图 20) ResNet-152 的 GPU 利用率和存储带宽。 .....	vi
图 21) Inception-v3 的 GPU 利用率和存储带宽。 .....	vi

## 1 内容提要

本文档包含技术白皮书 [WP-7267：可扩展的 AI 基础架构](#) 中所述架构的验证信息。 [Scalable AI Infrastructure](#)。该白皮书中所述设计使用 [NetApp® AFF A800](#)（一种全闪存 FAS 系统）、[NVIDIA® DGX-1™](#) 服务器和 [Cisco® Nexus® 3232C](#) 100 Gb 以太网交换机实施。我们使用行业标准基准测试工具验证了此系统的运行情况和性能，从验证测试结果来看，此架构可提供出色的训练和推理性能。验证结果同时也证明，存储性能余量足以支持多台 DGX-1 服务器。您还可以轻松、独立地将计算和存储资源从半机架配置扩展到多机架配置，并凭借可预测的性能满足任何机器学习工作负载要求。

## 2 计划摘要

### 2.1 经验证的 NetApp 架构计划

经验证的 NetApp 架构计划为客户提供适用于 NetApp 解决方案的经过验证的架构。借助经验证的 NetApp 架构，您可以获得具有以下优势的 NetApp 解决方案架构：

- 已经过全面测试
- 规范性
- 尽可能降低部署风险
- 加快上市速度

本文档主要面向 NetApp 及其合作伙伴的解决方案工程师以及客户的战略决策者。文档介绍了用于确定特定环境中所需的具体设备、布线和配置的架构设计注意事项。

### 2.2 NetApp ONTAP AI 解决方案

NetApp ONTAP® AI 融合基础架构由 NVIDIA DGX-1 服务器和 NetApp 云互联存储系统提供支持，是 NetApp 与 NVIDIA 联合开发和验证的架构。它将为您的企业提供具有以下优势的规范化架构：

- 消除复杂设计
- 支持独立扩展计算和存储
- 支持从小规模起步，然后无缝扩展
- 提供广泛的存储选项，满足各种性价比需求

NetApp ONTAP AI 将 NVIDIA DGX-1 服务器、NVIDIA Tesla® V100 图形处理单元 (GPU) 和 NetApp AFF A800 系统与一流的网络相集成。NetApp ONTAP AI 可消除复杂设计，避免盲目猜测，从而简化人工智能 (AI) 的部署。贵企业可以从小规模起步然后进行无中断扩展，同时还能智能地管理从边缘到核心再到云以及反向的数据传输。

图 1 显示了 NetApp ONTAP AI 解决方案的可扩展性。AFF A800 系统使用 4 台 DGX-1 服务器进行了验证，结果证明其具有足够的性能余量，可支持 5 台或更多 DGX-1 服务器，但不会影响存储吞吐量或延迟。此外，通过在 ONTAP 集群中添加更多网络交换机和存储控制器对，该解决方案可扩展到多个机架，提供极高吞吐量，从而加快训练和推理过程。采用这种方法时，可以根据数据湖大小、所使用深度学习 (DL) 模型以及所需性能指标灵活地独立调整计算与存储的比例。

图 1) NetApp ONTAP AI 解决方案机架级架构。



每个机架中 DGX-1 服务器和 AFF 系统的数量取决于所使用机架的电源和散热规格。系统的最终位置取决于计算流体动力学分析、气流控制 and 数据中心设计。

### 3 深度学习数据管道

深度学习是一个引擎，可帮助您检测欺诈行为、改善客户关系，优化供应链、交付创新产品和服务，从而在竞争日益激烈的市场中占据一席之地。深度学习模型的性能和准确性可通过扩大神经网络的规模和复杂性以及提高训练模型所使用数据的数据量和质量而得到大幅改进。

在具备海量数据集的情况下，搭建一个能够让您灵活地跨环境部署的基础架构是非常重要的。概括地讲，端到端深度学习部署由数据传输所经历的三个阶段组成：边缘（数据载入）、核心（训练集群和数据湖）和云（归档、分层和开发/测试）。这在数据跨数据管道所有三个阶段的物联网 (IoT) 等应用中非常典型。

图 2) 从边缘到核心再到云端的数据管道。



图 2 简要介绍了每个阶段的组成要素：

- **数据载入。**数据载入通常发生在边缘，例如，从自动驾驶汽车或销售点 (POS) 设备上捕获数据流。根据使用情形，可能需要在载入点或其附近部署 IT 基础架构。例如，零售商可能需要在每个商店占用少量空间来整合多台设备中的数据。
- **数据预处理。**预处理是在训练前对数据进行清理使其规范化的必要步骤。预处理发生在数据湖中，数据湖可能是以 Amazon S3 层形式存在于云环境中，也可能以文件存储或对象存储形式存在于内部存储系统中。



- **训练。**在关键的深度学习训练阶段，通常会定期将数据从数据湖复制到训练集群中。此阶段所用服务器使用 GPU 并行进行计算，从而形成巨大的数据处理能力。满足原始 I/O 带宽需求对于保持高 GPU 利用率至关重要。
- **推理。**经过训练的模型在测试后再部署到生产环境中。或者，可以将它们馈送回数据湖，以进一步调整输入权重；亦或在 IoT 应用中，也可以将这些模型部署到智能边缘设备。
- **归档、分层。**来自过去迭代的冷数据可能会无限期保存。许多人工智能团队倾向于将冷数据以对象存储形式归档到私有云或公共云中。

根据应用程序的不同，深度学习模型将处理大量不同类型的数据（结构化数据和非结构化数据）。这种不同会对底层存储系统提出各种不同的要求，包括存储数据的大小以及数据集中文件的数量。

其中一些高标准存储要求如下：

- 能同时存储和检索数百万个文件
- 存储和检索多种数据对象，例如图像、音频、视频和时序数据
- 以低延迟提供并行高性能以满足 GPU 处理速度要求
- 实现跨边缘、核心和云的无缝数据管理和数据服务

AFF 系统与 NetApp ONTAP 的卓越云集成功能以及软件定义的功能相结合，可支持跨边缘、核心和云的完整深度学习数据管道。本文档重点介绍数据管道的训练和推理组件的解决方案。

## 4 解决方案概述

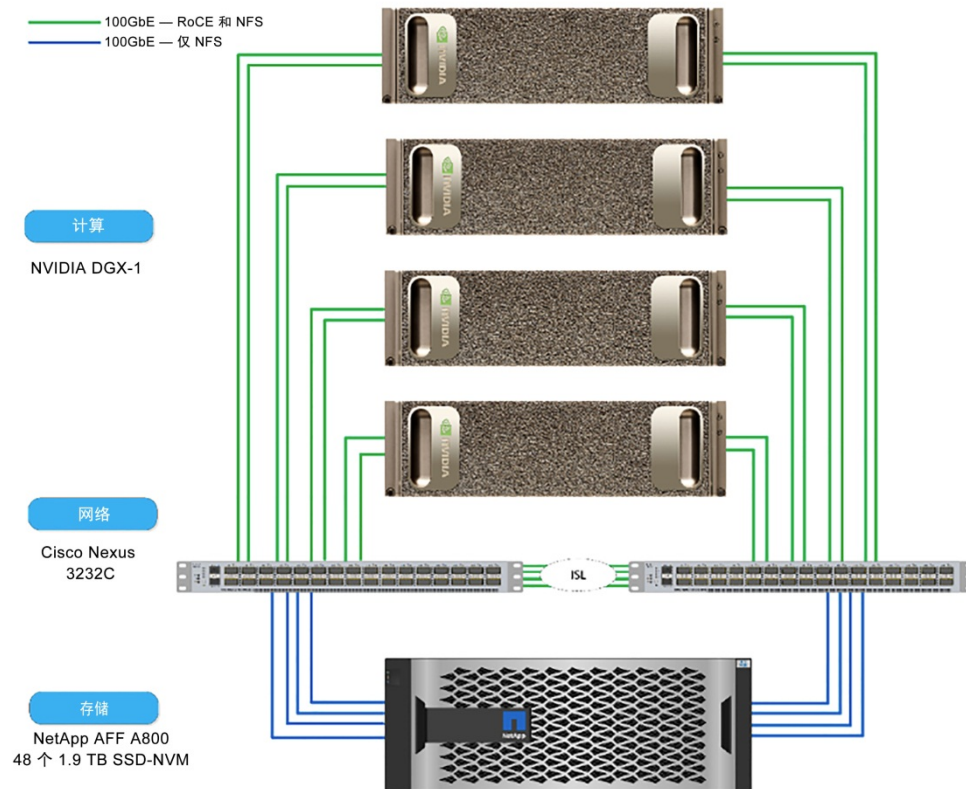
深度学习系统所采用的算法属于计算密集型算法，而且特别适合 NVIDIA GPU 的架构。深度学习算法中执行的计算涉及并行运行的巨量矩阵乘法。相较于通用中央处理单元 (CPU)，现代化 GPU 的高度并行架构可大幅提高深度学习等并行处理数据的应用程序的效率。利用 DGX-1 服务器的单个和集群 NVIDIA GPU 计算架构的发展进步使其成为高性能计算 (HPC)、深度学习和分析等工作负载的首选平台。要在这些环境中将性能最大化，需要一个能够不断为 NVIDIA GPU 馈送数据的支持基础架构。因此，必须能以超低延迟和高带宽访问数据集。

### 4.1 解决方案技术

此解决方案采用 1 个 NetApp AFF A800 系统、4 台 NVIDIA DGX-1 服务器和 2 个 Cisco Nexus 3232C 100 Gb 以太网交换机实施。每台 DGX-1 服务器均使用基于融合以太网的远程直接内存访问 (RDMA) (RDMA over Converged Ethernet, RoCE)，通过用于 GPU 间通信的 4 个 100GbE 连接与 Nexus 交换机相连。用于 NFS 存储访问的传统 IP 通信也发生在这些链路上。每个存储控制器均使用 4 个 100GbE 链路连接到网络交换机。

因为基于 InfiniBand (IB) 的 RDMA 具有高带宽和低延迟特性，所以传统 HPC 基础架构将其用于节点间连接。随着以太网技术达到以前只有使用 IB 才可能达到的性能级别，再加上以太网技术得到充分的了解并且广泛部署于每个企业数据中心，因此 RoCE 使得这些功能的采用变得更加简单。图 3 显示了基础解决方案架构。

图 3) 经验证的 NetApp ONTAP AI 解决方案架构。



## 4.2 NVIDIA DGX-1 服务器

DGX-1 服务器是一个专门为深度学习工作流程构建且软硬件全面集成的统包系统。每台 DGX-1 服务器均由 8 个 Tesla V100 GPU 提供支持，配置在使用 NVIDIA NVLink™ 技术的混合式立方体网络拓扑结构中，可为 GPU 间通信提供超高带宽、低延迟网络结构。这种拓扑结构对于多 GPU 训练必不可少，消除了基于 PCIe 的互连无法随 GPU 数量增长而实现性能线性提升的瓶颈。DGX-1 服务器还配备了高带宽、低延迟网络互连，用于在支持 RDMA 的网络结构上实现多节点集群。

DGX-1 由 NVIDIA GPU Cloud (NGC) 提供支持，NGC 是 NVIDIA 用于 GPU 加速软件、基于云的容器注册表。NGC 为当今针对 NVIDIA GPU 进行了优化的最常见深度学习框架（例如 Caffe2、TensorFlow、PyTorch、MXNet 和 TensorRT 等）提供容器。这些容器集成了框架或应用程序、必需驱动程序、库和通信原语，并由 NVIDIA 在整个堆栈上进行了优化，可提供最高 GPU 加速性能。NGC 容器采用 NVIDIA CUDA 工具包，其中包含 NVIDIA CUDA 基础线性代数子程序库 (CUDA Basic Linear Algebra Subroutines Library, cuBLAS)、NVIDIA CUDA 深度神经网络库 (CUDA Deep Neural Network Library, cuDNN) 等。NGC 容器还包含用于多 GPU 和多节点总体通信原语的 NVIDIA 集体通信库 (NVIDIA Collective Communications Library, NCCL)，从而在进行深度学习训练时建立拓扑结构意识。NCCL 可以在一台 DGX-1 服务器内部以及多台 DGX-1 服务器之间实现 GPU 间通信。

## 4.3 NetApp AFF 系统

NetApp AFF 是一款一流的存储系统，具有行业领先的性能、卓越的灵活性、云集成及同类最佳数据管理功能，可帮助您满足企业级存储需求。AFF 系统专为闪存设计，有助于加速、管理和保护业务关键型数据。

NetApp AFF A800 系统是行业首款端到端 NVMe 解决方案。对于 NAS 工作负载而言，一个 AFF A800 系统支持吞吐量为 25 GB/秒的顺序读取和 100 万次 IOPS 的小型随机读取，同时保持低于 500 微秒的延迟。AFF A800 系统支持以下功能：

- 在包含 24 个节点的集群中实现最高 300 GB/秒的巨大吞吐量和 1140 万次 IOPS
- 100 Gb 以太网连接与 32 Gb FC 连接相结合
- 支持多流写入 (MSW) 的 30 TB 固态驱动器 (SSD)

- 在 2U 驱动器架中实现 2 PB 的高密度容量
- 从 364 TB（2 个节点）扩展到 74 PB（24 个节点）
- NetApp ONTAP 9.4，具有一整套数据保护和复制功能，可提供行业领先的数据管理

从性能上来说，AFF A700s 属于下一代最佳存储系统，可以为 NAS 工作负载提供 18 GB/秒的吞吐量和 40GbE 的传输速度。AFF A300 和 AFF A220 系统以更低的成本提供充足的性能。

## 4.4 NetApp ONTAP 9

ONTAP 9 是 NetApp 推出的最新一代存储管理软件，可帮助您打造现代化基础架构并向云就绪数据中心过渡。利用行业领先的数据管理功能，无论数据驻留于何处，ONTAP 均支持您使用同一套工具管理和保护数据。还可以自由地将数据迁移到任何需要的地方，无论是边缘、核心还是云端。ONTAP 9 包含许多功能，不仅可以跨不同混合云架构简化数据管理，加速提供并保护关键数据，而且能帮助打造适应未来需求的基础架构。

### 简化数据管理

数据管理对于企业 IT 运营至关重要，通过得当的管理才能将适当的资源用于应用程序和数据集。ONTAP 包含以下可简化运营和降低 TCO 的功能：

- **实时数据缩减和扩展的重复数据删除。**数据缩减可减少存储块内的空间浪费，而重复数据删除可大幅提升有效容量。
- **最低、最高和自适应服务质量 (QoS)。**精细 QoS 控制可帮助您在高度共享环境中保持关键应用程序的性能级别。
- **ONTAP FabricPool。**此功能支持将冷数据自动分层到公共云和私有云存储中，其中包括 Amazon Web Services (AWS)、Azure 和 NetApp StorageGRID® 解决方案。

### 加快数据访问速度并提供数据保护

ONTAP 可提供卓越的性能和数据保护，而且可通过以下方式扩展这些功能：

- **高性能和低延迟。**ONTAP 尽可能以最低的延迟提供最高的吞吐量。
- **NetApp ONTAP FlexGroup。**FlexGroup 卷是高性能数据容器，可线性扩展到多达 20 PB，能容纳 4000 亿个文件，提供了一个可简化数据管理的单一命名空间。
- **数据保护。**ONTAP 提供适用于所有平台的内置数据保护功能和通用管理功能。
- **NetApp 卷加密。**ONTAP 提供同时支持板载和外部密钥管理的本机卷级加密。

### 适应未来需求的基础架构

ONTAP 9 可帮助您满足瞬息万变的严苛业务需求：

- **无缝扩展且无中断运行。**ONTAP 支持向现有控制器和横向扩展集群无中断添加容量。您可以升级到 NVMe 和 32 Gb FC 等最新技术，而无需进行代价高昂的数据迁移或中断。
- **云连接。**ONTAP 是云互联支持最广泛的存储管理软件，在所有公共云中均提供适用于软件定义的存储 (ONTAP Select) 和云原生实例 (NetApp Cloud Volumes Service) 的选项。
- **与新兴应用程序相集成。**ONTAP 使用支持现有企业级应用程序的相同基础架构，为 OpenStack、Hadoop 和 MongoDB 等下一代平台和应用程序提供企业级数据服务。

## 4.5 NetApp FlexGroup 卷

训练数据集通常是由大量（可能高达数十亿）文件组成的集合。此类文件可能包括文本、音频、视频以及其他形式的非结构化数据，必须进行存储和处理才能并行读取。存储系统必须存储大量小文件（可能高达数十亿），而且必须并行读取这些文件才能处理顺序和随机 I/O。

FlexGroup 卷（图 4）是由多个成员卷组成的单一命名空间，对存储管理员而言，就像 NetApp FlexVol® 卷一样加以管理和使用。FlexGroup 卷中的文件分配给各个成员卷，而且不会跨卷或节点进行条带化。它们支持以下功能：



- FlexGroup 卷可为拥有庞大元数据的工作负载提供巨大容量（数 PB）和可预测的低延迟。
- 它们支持在同一个命名空间中存储数千亿个文件。
- 它们支持跨 CPU、节点、聚合和成员 FlexVol 卷并行运行 NAS 工作负载。

图 4) NetApp FlexGroup 卷。



#### 4.6 NVIDIA GPU Cloud 和 Trident

NVIDIA GPU Cloud (NGC) 为深度学习充分利用 NVIDIA GPU 提供了一系列经过性能设计的全面集成 Docker 映像。这些映像包括 NVIDIA CUDA 工具包和 NVIDIA 深度学习库等所有必需的依赖项。这些映像经过 NVIDIA 测试、优化和认证，可用于 NVIDIA DGX-1 服务器。而且，为了让利用 GPU 的映像变得可移植，NVIDIA 开发了 NVIDIA Container Runtime for Docker，支持您在启动时将 NVIDIA 驱动程序和 GPU 的用户模式组件挂载到 Docker 容器中。

NetApp 提供的 Trident 是适用于 Docker 和 Kubernetes 的开源动态存储配置程序。Trident 与 NGC 和 Kubernetes 或 Docker Swarm 等常见业务流程协调程序相结合，支持您将深度学习 NGC 容器映像无缝部署到 NetApp 存储上，从而获得企业级人工智能容器部署体验。此类部署包括自动化流程编排、用于测试和开发的克隆、使用克隆进行升级测试、用于保护和满足合规性要求的副本以及针对 NGC AI 和 DL 容器映像的更多数据管理用例。

#### 4.7 Cisco Nexus 3232C 网络交换机

Cisco Nexus 3232C 交换机（图 5）是专为数据中心设计的低延迟、高密度、高性能、节能型 100 Gb/秒交换机。这种紧凑型 1 机架单元 (1RU) 型号能在所有端口上以 450 纳秒延迟提供线速第 2 层和第 3 层交换。这款交换机是 Cisco Nexus 3200 平台的成员，运行行业领先的 Cisco NX-OS 软件操作系统，可为您提供全面且部署广泛的特性和功能。Cisco Nexus 3232C 是带有 32 个 QSFP28 端口的四通道小型可热插拔 (Quad Small Form-Factor Pluggable, QSFP) 交换机。每个 QSFP28 端口均能以 10、25、40、50 和 100 Gb/秒的速度运行，25 Gb/秒的最大端口数可达 128 个。

图 5) 为融合增强型以太网标准以及 RoCE v1 和 v2 提供 NX-OS 支持的 Cisco Nexus 交换机。



根据测试，此解决方案仅使用每个网络交换机上一半的可用端口。每个交换机最多可支持 8 台 DGX-1 服务器，并通过附加存储访问端口提供更多 GPU 处理能力。对于更大规模的实施，Cisco Nexus 7000 的每个交换机最多可支持 192 个 100GbE 端口。或者，也可以使用连接到中央主干交换机的多对 Nexus 3000 交换机实施分支-主干拓扑结构。

## 4.8 基于融合以太网的 RDMA

通过直接内存访问 (DMA)，磁盘驱动器控制器、声卡、显卡和网卡等硬件子系统可以在不使用 CPU 处理周期的情况下，访问系统内存来执行数据读/写操作。RDMA 进一步扩展了这种能力，它通过使用零复制功能，允许网络适配器在应用程序内存之间进行服务器到服务器的数据传输，而不涉及任何操作系统或设备驱动程序。这种方法通过绕过内核进行读/写和发送/接收操作，大幅降低了 CPU 开销和延迟。

就实施基于以太网的 RDMA 而言，RoCE 部署范围最为广泛，而且它采用了全新的融合增强型以太网 (Converged Enhanced Ethernet, CEE) 标准。现在，它在许多高端网络适配器、融合网络适配器和网络交换机中已成为标配功能。传统以太网对网络流量采用尽力确保最佳结果的交付机制，对于需要低延迟和高带宽的 GPU 节点间通信而言并不适合。CEE 支持无损物理层网络介质，并且能够选择性地为网络上的任何特定流量分配带宽。

为了保证无损、有序地交付以太网数据包，CEE 网络使用优先级流量控制 (Priority Flow Control, PFC) 和增强传输选择 (Enhanced Transmission Selection, ETS)。PFC 支持为每个特定服务类别 (Class of Service, CoS) 发送暂停帧，因此您可以在限制特定网络流量的同时允许其他流量自由流动。ETS 支持为每个 CoS 分配特定带宽，以便更精细地控制网络利用率。

由于能让 RoCE 优先于其他所有流量，因此可将 100GbE 链路同时用于传输 RoCE 流量和传统 IP 流量，例如此解决方案中展示的 NFS 存储访问流量。

## 5 技术要求

本节介绍验证此解决方案时使用的硬件和软件。第 7 节解决方案验证记录的所有测试均为使用本节所列硬件和软件执行。

**注：**此参考架构中验证的配置均基于可用的实验室设备，而非基于对所测试硬件的要求或限制。

### 5.1 硬件要求

表 1 列出了验证此解决方案所使用的硬件组件。根据您的要求，具体实施此解决方案时所使用的硬件组件可能会有所不同。

表 1) 硬件要求。

硬件	数量
NVIDIA DGX-1 GPU 服务器	4
NetApp AFF A800 系统	1 个高可用性 (HA) 对，包括 48 个 1.92 TB NVMe SSD
Cisco Nexus 3232C 网络交换机	2

### 5.2 软件要求

表 2 列出了实施此解决方案所需的软件组件。根据您的要求，具体实施此解决方案时所使用的软件组件可能会有所不同。

表 2) 软件要求。

软件	版本
NetApp ONTAP	9.4
Cisco NX-OS 交换机固件	7.0(3)I6(1)
NVIDIA DGX-1 操作系统	Ubuntu 16.04 LTS
Docker 容器平台	18.03.1-ce [9ee9f40]
容器版本	基于 nvcr.io/nvidia/tensorflow:18.04-py2 的 netapp_1.7.0.2

软件	版本
机器学习框架	TensorFlow 1.7.0
Horovod	0.11.3
OpenMPI	3.1.0
基准测试软件	TensorFlow 基准测试 [1b1ca8a]

## 6 解决方案架构

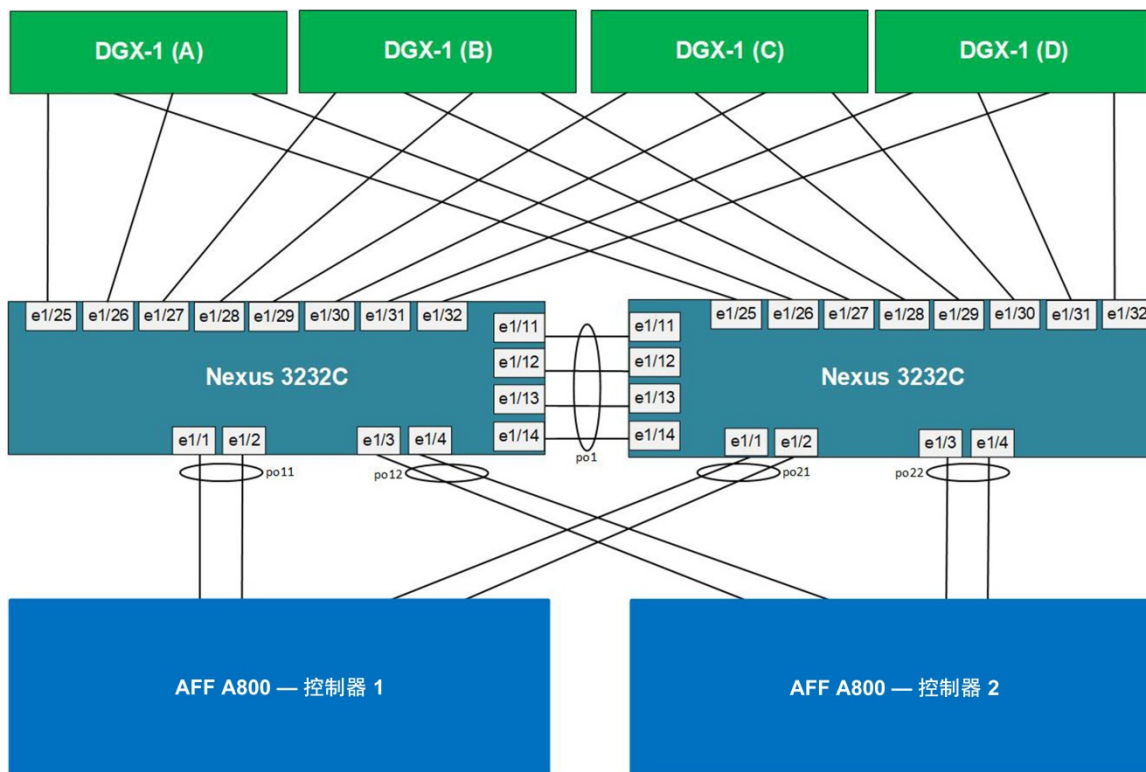
此架构经过验证，可满足运行深度学习工作负载的要求。有了这个验证结果，数据科学家可以在经过预先验证的基础架构上部署深度学习框架和应用程序，因此有助于消除风险，让企业集中精力从数据中获得有价值的洞察。此架构还可以为其他 HPC 工作负载提供出色的存储性能，而且无需对基础架构进行任何修改或调整。

### 6.1 网络拓扑结构和交换机配置

在此解决方案中，RoCE 用来代替 IB 提供 DGX-1 服务器间通信所需的高带宽、低延迟连接。Cisco Nexus 交换机通过实施 PFC 来支持 RoCE，这样用户可以在共享链路上让 RoCE 流量优先于传统 IP 流量，并且允许将 100GbE 链路同时用于 RoCE 和 IP。

此架构使用一对 Cisco Nexus 3232C 100 Gb 以太网交换机来构建集群间和存储访问主网络。这两个交换机使用配置为标准端口通道的四个 100 Gb 网络端口互连。在主机或存储系统链路发生故障时，该交换机间链路 (Inter Switch Link, ISL) 端口通道允许流量在交换机之间流动。每个主机均通过一对“活动-被动”绑定连接到 Nexus 交换机，而且为了提供链路层冗余，每个存储控制器均通过双端口 LACP 端口通道连接到每个 Nexus 交换机。图 6 显示了网络交换机端口配置。

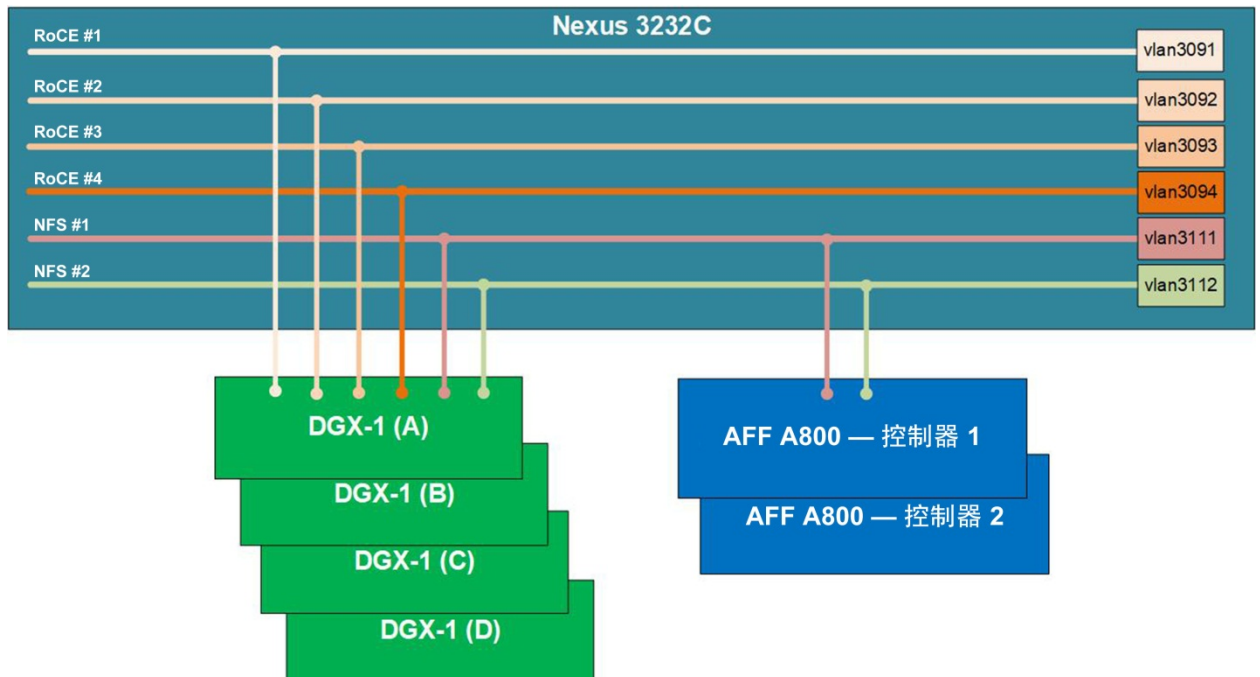
图 6) 网络交换机端口配置。



配置了多个虚拟 LAN (VLAN) 来同时支持 RoCE 和 NFS 存储流量。四个 VLAN 专用于 RoCE 流量，两个 VLAN 专用于 NFS 存储流量。四个不同 VLAN 和 IP 范围用于为 RoCE 连接提供对称路由，NVIDIA 软件堆栈管理这些连接以实现带宽聚合和容错功能。对于存储访问，此解决方案使用并不支持多路径访问的 NFSv3，因此使用两个 VLAN 来支持多个专用 NFS 挂载。此方法虽不提供任何额外容错功能，但确实支持使用多个链路增加可用带宽。每个交换机上均配置了 PFC，所有四个 RoCE VLAN 分配给优先类别，NFS VLAN 分配给默认尽力确保最佳结果的类别。所有 VLAN 均配置为可供巨型帧使用，最大传输单元 (Maximum Transmission Unit, MTU) 大小设置为 9000。

DGX-1 服务器的交换机端口被配置为中继端口，所有 RoCE VLAN 和 NFS VLAN 均可使用。为存储系统控制器配置的端口通道虽然也是中继端口，但只允许 NFS VLAN 使用。图 7 显示了 DGX-1 服务器和存储系统端口的 VLAN 连接。

图 7) DGX-1 和存储系统端口的 VLAN 连接。



为了优先为 RoCE 流量提供服务，主机网络适配器为每个 RoCE VLAN 上的流量分配的 CoS 值为 4。为交换机配置的 QoS 策略是为具有此 CoS 值的流量提供无丢包的服务。为 NFS 流量分配的默认 CoS 值为 0，表示此类流量适用交换机上的默认 QoS 策略，所获服务为尽力确保最佳结果服务。

然后在每个 DGX-1 端口上启用 PFC，这样交换机端口就可以为特定服务类别发送暂停帧，以避免交换机发生拥塞。发生拥塞时，这种配置通过使用 ETS 将 95% 的带宽分配给 RoCE 流量，可以在 RoCE 与 NFS 流量之间动态分配资源，同时优先满足节点到节点的通信。您也可以动态修改带宽分配，以便针对需要更高存储性能但节点间的通信较少的工作负载进行优化。

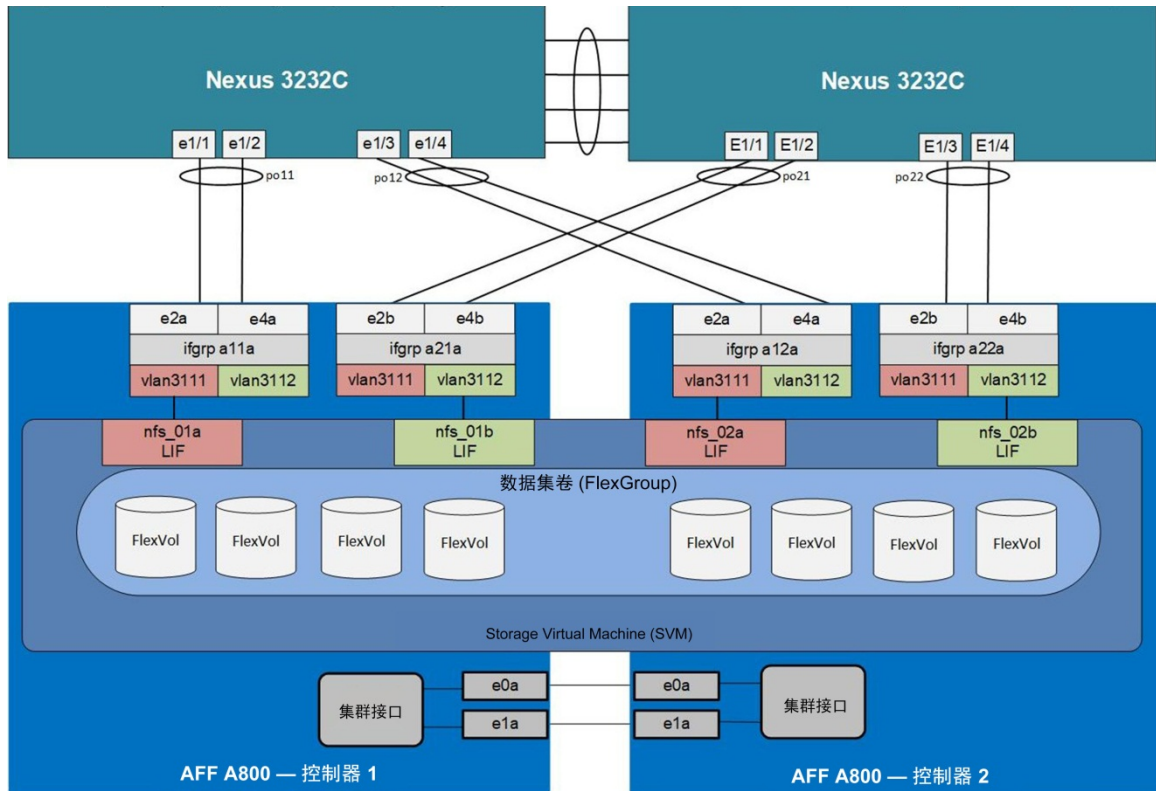
## 6.2 存储系统配置

为了满足此架构中任何潜在工作负载的存储网络需求，除了存储集群互连所需的板载端口外，每个存储控制器还配置了四个 100GbE 端口。图 8 显示了存储系统配置。在每个控制器上为每个交换机配置了一个双端口 LACP 接口组（图 8 中的 ifgrp）。这些接口组可提供与每个交换机高达 200 Gb/秒的弹性连接以用于数据访问。为 NFS 存储访问配置了两个 VLAN，两个存储 VLAN 都从交换机中继到这些接口组中的每个接口组。这种配置支持通过多个接口从每个主机并发访问数据，这样可增加可用于每个主机的潜在带宽。

从存储系统进行的所有数据访问均通过从专用于此工作负载的 Storage Virtual Machine (SVM) 的 NFS 访问提供。该 SVM 配置了总共四个逻辑接口 (LIF)，其中每个存储 VLAN 上两个 LIF。每个接口组托管一个 LIF，这样即是每个控制器上的每个 VLAN 一个 LIF，每个 VLAN 一个专用接口组。不过，两个 VLAN 均被中继到每个控制器上的两个接口组。此配置支持每个 LIF 故障转移到同一个控制器上的另一个接口组，这样两个控制器在发生网络故障时均可保持活动状态。



图 8) 存储系统配置。



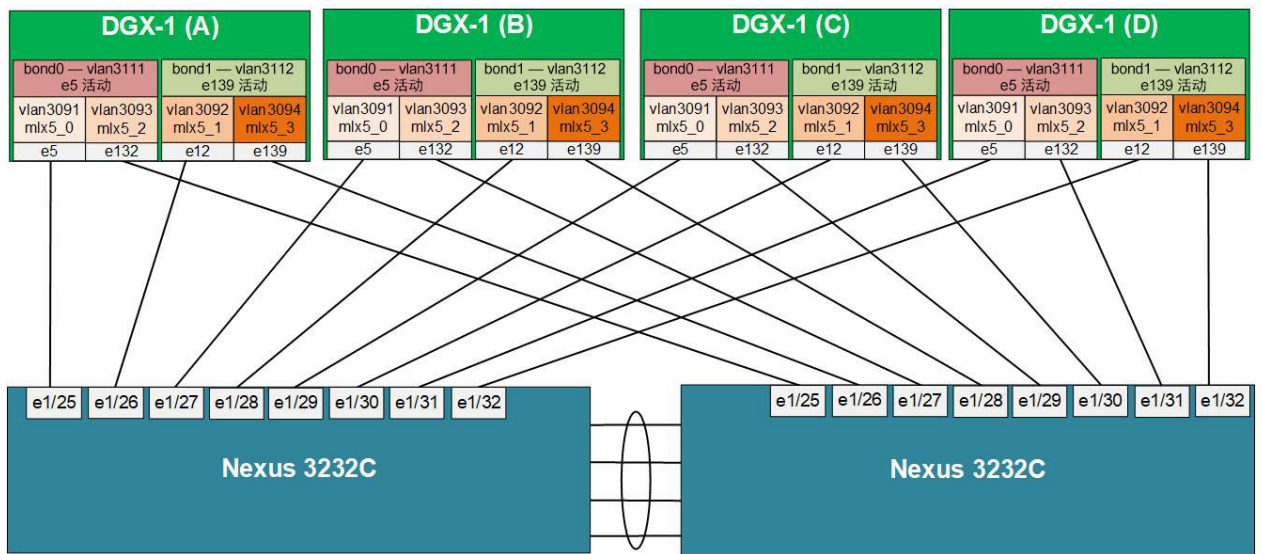
对于逻辑存储配置，此解决方案使用 FlexGroup 卷提供一个单一的存储池，该存储池分布在存储集群中的各节点上。每个控制器均托管一个含 46 个磁盘分区的聚合，每个磁盘由两个控制器共享。在数据 SVM 上部署 FlexGroup 时，便会在每个聚合上配置许多 FlexVol 卷，这些卷再结合成为 FlexGroup。通过这种方式，存储系统可提供单一存储池，其容量最高可扩展到阵列的最大容量，而且可通过同时利用阵列中的所有 SSD 提供出色的性能。NFS 客户端可通过为该 SVM 配置的任何 LIF 访问作为单个挂载点的 FlexGroup。您只需在存储集群中添加更多节点即可增加容量和客户端访问带宽。

### 6.3 主机配置

对于网络连接，每个 DGX-1 均配置了四个 Mellanox ConnectX4 单端口网络接口卡。这些网络接口卡运行速度最高可达 100 Gb 以太网速度，而且支持 RoCE，从而为集群互连应用程序所使用的 IB 提供成本更低的替代方案。每个 100 Gb 端口均在相应的交换机上配置为中继端口，每个交换机上允许配置四个 RoCE VLAN 和两个 NFS VLAN。图 9 显示了 DGX-1 主机的网络端口和 VLAN 配置。



图 9) DGX-1 主机的网络端口和 VLAN 配置。



对于 RoCE 连接，每个物理端口均在四个 RoCE VLAN 中的其中一个上托管一个 VLAN 接口和 IP 地址。Mellanox 驱动程序配置为将网络 CoS 值 4 应用于每个 RoCE VLAN，而 PFC 在交换机上配置为保证优先为 RoCE 类别提供无损服务。RoCE 虽不支持将多个链路聚合为一个逻辑连接，但 NVIDIA NCCL 通信软件可以使用多个链路实现带宽聚合和容错功能。

对于 NFS 存储访问，通过使用与每个交换机的一个链路创建两个“主动-被动”绑定。每个绑定将一个 VLAN 接口和 IP 地址托管在两个 NFS VLAN 的其中一个上，而且每个绑定的主动端口连接到不同的交换机。此配置可在每个 NFS VLAN 中提供最高 100 Gb 的带宽，并且可以在发生任何主机链路或交换机故障时提供冗余。为了给 RoCE 连接提供最佳性能，所有 NFS 流量均被分配给默认尽力保证最佳效果的 QoS 类别。所有物理接口和绑定接口的 MTU 均配置为 9000。

为了提高数据访问性能，从 DGX-1 服务器到存储系统进行了多个 NFSv3 挂载。每个 DGX-1 服务器均配置了两个 NFS VLAN，每个 VLAN 上一个 IP 接口。AFF A800 系统上的 FlexGroup 卷挂载在每个 DGX-1 中的每个 VLAN 上，从而提供从服务器到存储系统的完全独立的连接。虽然一个 NFS 挂载也能提供此工作负载所需的性能，但若定义多个挂载点，便可将更多存储访问带宽用于存储需求更大的其他工作负载。

## 7 解决方案验证

本节介绍我们为验证此解决方案的运行和性能而执行的测试。我们使用第 5 节技术要求所列的特定设备和软件执行了本节所述的全部测试。

### 7.1 验证测试计划

此解决方案使用包含各种计算配置的标准基准测试进行了验证，结果证明了这种架构的可扩展性。ImageNet 数据集使用一个 FlexGroup 卷托管在 AFF A800 系统上，按照 NVIDIA 对外部存储访问的建议，该卷由最多四个 DGX-1 服务器使用 NFSv3 进行访问。TensorFlow 作为所测试全部模型的机器学习框架，每个测试案例均采集了计算和存储性能指标。第 7.2 节验证测试结果列出了这些数据的亮点。

下列具有不同计算和存储复杂程度的卷积神经网络 (Convolutional Neural Network, CNN) 模型被用来展现训练速度：

- **ResNet-152** 通常被视为最准确的训练模型。
- **ResNet-50** 在处理速度更快时准确性高于 AlexNet。
- **VGG16** 可产生最高的 GPU 间通信需求。
- **Inception-v3** 是另一个通用 TensorFlow 模型。

上述每个模型均使用各种不同的硬件和软件配置进行了测试，用于研究每种方案对性能的影响：

- 我们使用合成数据和 ImageNet 参考数据集对每个模型进行了测试。我们使用更多 GPU 在 DGX-1 服务器内部以及多个 DGX-1 服务器之间进行了进一步测试，帮助评估了计算集群的可扩展性以及存储访问的性能。
- 我们在将数据复制到 GPU 内存中之前，使用禁用了失真的 ImageNet 数据降低了 CPU 处理的开销。
- 我们使用 Tensor 核心和 CUDA 核心测试了每个模型，证明了 Tensor 核心能够提升性能。
- 提高 GPU 性能还具有提高存储访问需求的作用，证明 AFF A800 系统能够轻松满足这些需求。
- 我们使用不同的批大小测试了每个深度学习模型。增加批大小会对系统产生多项影响，最终会导致总训练速度提高，GPU 间通信需求降低以及存储带宽需求增加。我们对每个模型进行了下列批大小的测试：
  - 对 ResNet-50 使用了 64、128 和 256
  - 对其他所有模型使用了 64 和 128
- 每个模型分别使用 1 个、2 个和 4 个 DGX-1 服务器进行了测试，证明每个模型在使用 RoCE 作为互连方式（通过 Horovod）的多个 GPU 之间可以实现扩展。
- 我们使用所有模型并采用最大批大小（ResNet-50 使用 256，其他所有模型使用 128）、32 个 GPU（Tensor 核心和 CUDA 核心）和 ImageNet 数据集运行了推理。
- 在经过至少两次推理后，所有性能指标即收集完毕。在进行多次训练后，我们观察到性能略有提升。每个测试均进行五次，并报告我们观察到的性能指标的平均值。

## 7.2 验证测试结果

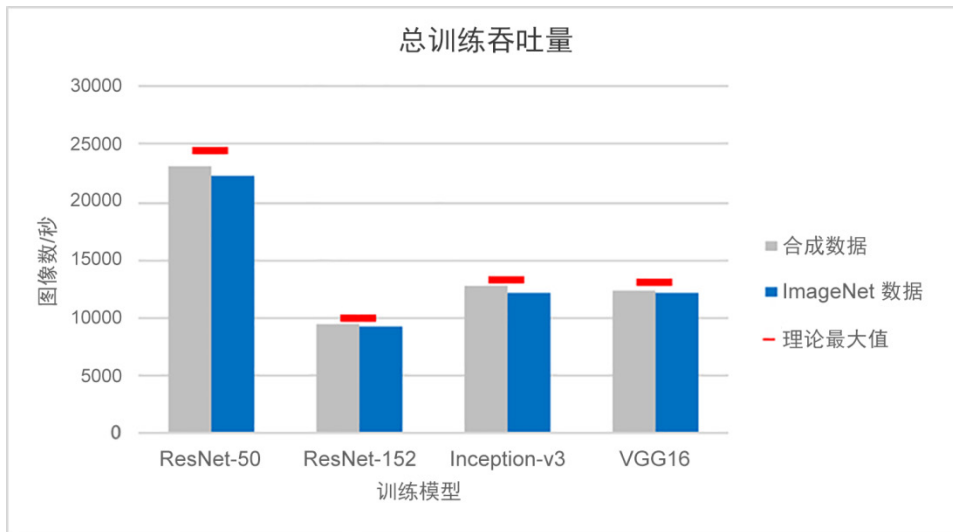
正如前文所述，为了评估此解决方案的总体运行情况和性能，我们进行了各种不同的测试。本节介绍在进行这些测试期间收集的計算和存储性能数据的亮点。完整的详细测试结果请参见附录。请注意以下有关本报告后续章节中所提供数据的详细信息：

- 模型训练性能按每秒的图像数来衡量。
- 存储性能使用吞吐量（MB/秒）和延迟（微秒）来衡量。另外还采集了存储系统 CPU 的性能数据来评估存储系统上的剩余性能容量。
- 每个系统均使用多个批大小进行了测试。较大的批大小会提高训练总吞吐量。此处仅列出了测试每个模型所用的最大批大小。所测试的每个批大小的数据请见附录：
  - ResNet-50 测试使用的批大小为 256。
  - ResNet-152、Inception-v3 和 VGG16 测试使用的批大小为 128。

### 总训练吞吐量

图 10 显示了所测试的每个模型使用 Tensor 核心达到最大性能时所实现的每秒最大训练图像数量。图 10 对使用 ImageNet 数据和合成数据进行基线比较时用 32 个 GPU 所达到的训练吞吐量进行了对比。图中同时还显示了可达到的理论最大值，该值为全部 GPU 独立训练合成数据而不相互更新参数时所达到的数量。如图 10 中所示，我们为 ImageNet 数据实现的吞吐量非常接近为合成数据实现的吞吐量。

图 10) 所有模型的训练吞吐量。

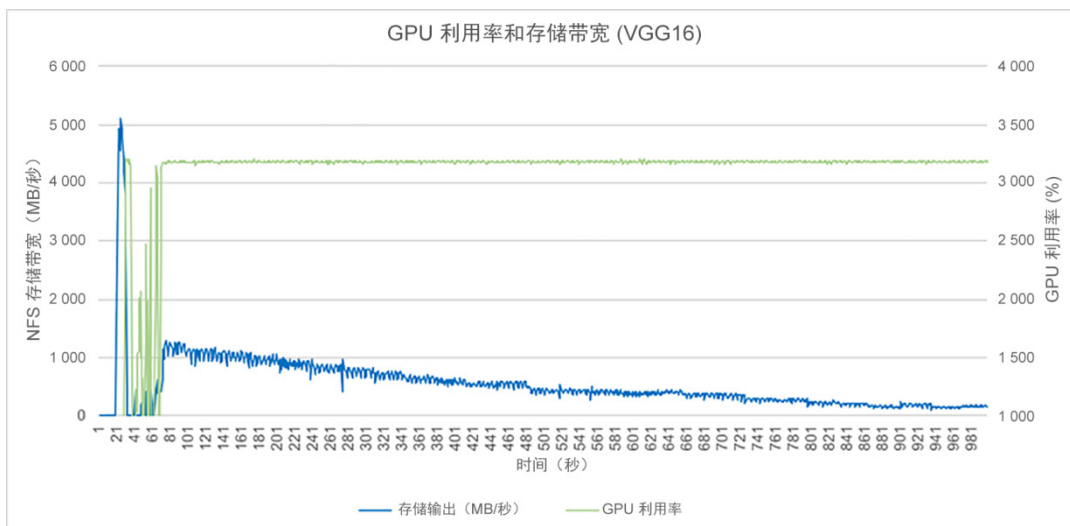


## GPU 工作负载性能

下一个数据集证明存储系统能够满足满负载 DGX-1 服务器的需求。图 11 显示了使用 32 个 GPU 运行每个模型时产生的 DGX-1 服务器的 GPU 利用率及存储带宽。从图中可以看到，存储带宽刚开始时非常高，因为需要将初始数据从存储读入 TensorFlow 管道缓存中，然后随着时间的推移，当大部分数据集驻留在 DGX-1 本地内存中时，存储带宽便逐渐降低。

在所有数据都位于本地内存中之后，存储访问下降到接近零。DGX-1 GPU 几乎立即开始处理数据，GPU 利用率在整个测试过程中始终保持平稳。此图显示了 VGG16 模型使用 128 批大小测试的结果，它在我们的测试中所产生的 GPU 利用率级别最高。其他模型的结果图表请参见附录。请注意，GPU 利用率数值为所有 GPU 的利用率之和，因此在使用 32 个 GPU 测试的本案例中，最大可能利用率为 3200%。

图 11) GPU 利用率和存储带宽 (VGG16)。



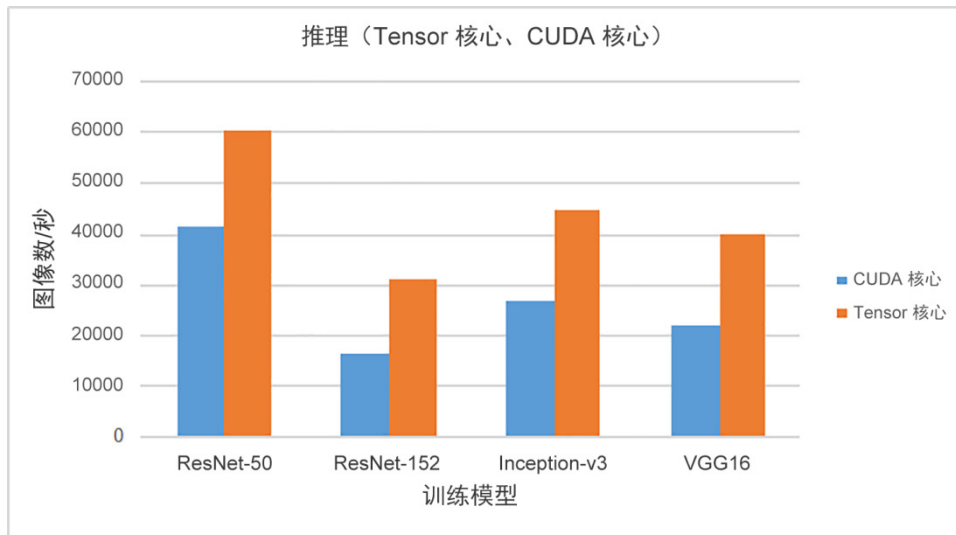
从图 11 中可以看到，所有 32 个 GPU 的利用率始终保持在 95% 以上，而且无论来自存储系统的数据量多大，它也始终保持平稳。存储系统初始数据吞吐量为 5 GB/秒，然后在训练过程的剩余时间里从 2 GB/秒左右降到接近零。这个结果证明，对于此工作负载而言，存储访问并非 GPU 性能的瓶颈。使用超出本地内存容量的更大规模数据集时，直到训练过程后期，存储访问性能将一直保持稳定的吞吐速率。此外，图 11 还比较了作为存储带宽函数的 GPU 利用率。它没有采集整个训练过程所需的时间，因为随着训练过程的进展，存储带宽逐渐下降到接近于零。

## 使用 GPU 进行推理

推理是部署深度学习模型来评估一系列新对象，并根据在训练过程中观察到的结果做出准确性相似的预测的过程。在处理图像数据集的应用程序中，推理的目标是将输入图像分类，然后尽可能迅速地请求者做出响应。除了实现高吞吐量外，尽可能降低延迟也很重要。

NetApp ONTAP AI 被用于证明推理，并衡量此过程的吞吐量指标。图 12 显示了在推理过程中每秒可以处理的图像数量。此测试比较了在使用 Tensor 核心和 CUDA 核心测试的每个模型上，使用 ImageNet 数据时用 32 个 GPU 所达到的吞吐量。借助 NetApp ONTAP AI 的处理能力，可以使用 Tensor 核心即时对数量庞大的图像进行分类。

图 12) 所有模型的推理（Tensor 核心和 CUDA 核心）。



## AFF A800 系统用于人工智能训练工作负载时的性能

所采集的存储带宽、延迟和 CPU 性能余量数据证明了存储系统用于所测试的每个模型时的性能。图 13 至图 15 显示了使用真实数据测试每个模型时的存储系统指标。这些围绕存储进行的测试使用更大的批大小进行，目的在于增加存储工作负载，展现最坏情况下的性能。

注意，在每个指标中，AFF A800 系统的性能都足以支持每个模型用 32 个 GPU 生成的总工作负载。为了提供训练工作负载的参考框架，使用具有 64K 顺序读取 I/O 配置文件的灵活 I/O (fio) 生成了一个人工工作负载。对于使用 fio 生成的工作负载，最大吞吐量超过了 15 GB/秒，而读取延迟始终保持在 1 毫秒以下，CPU 利用率略低于 50%。为了利用有限数量的可用 DGX-1 服务器实现可能的最大吞吐量，在每个服务器上使用了更多 NFS 挂载和多个 fio 作业。

**注：** 经证实，对于 NAS 工作负载，NetApp AFF A800 HA 对可在延迟低于 1 毫秒的情况下支持高达 25 GB/秒的吞吐量。

图 13) 所有模型的存储带宽。

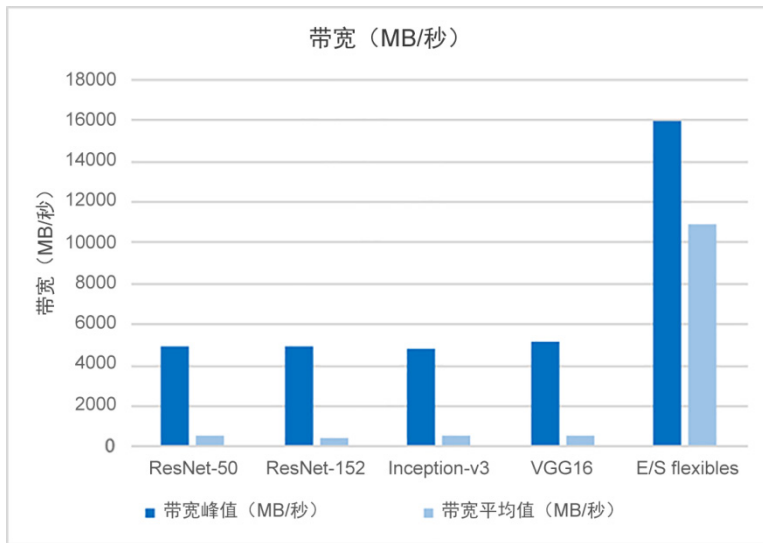


图 14) 所有模型的存储延迟。

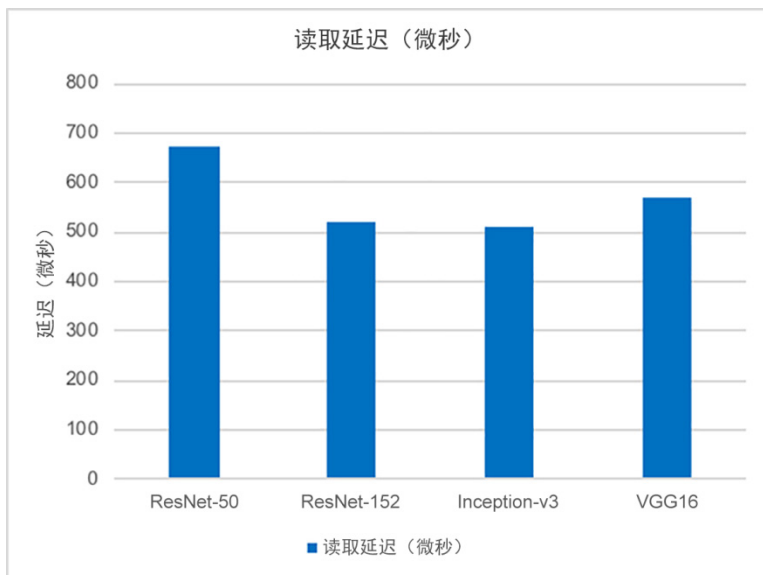
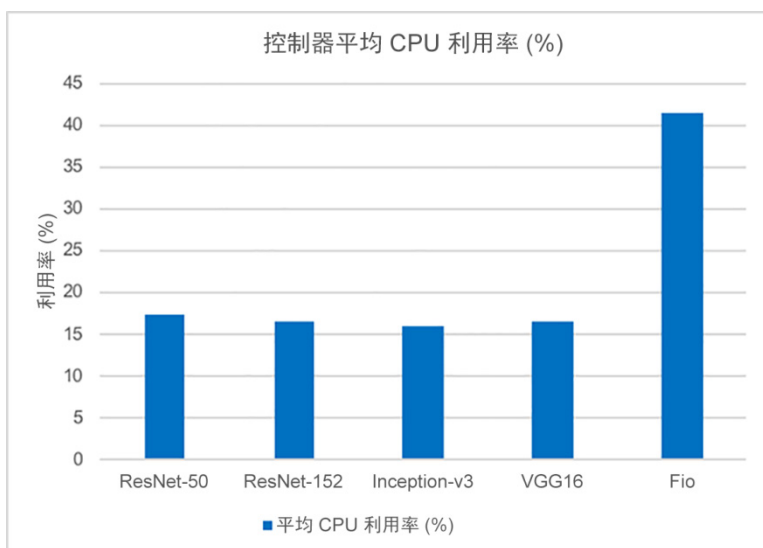


图 15) 所有模型的存储 CPU 利用率。





## 7.3 解决方案规模估算指导

此架构旨在供意欲采用 NVIDIA DGX-1 服务器和 NetApp AFF 系统实施高性能计算 (HPC) 基础架构的客户及合作伙伴作为参考。

此次验证证明, AFF A800 系统可轻松支持四个 DGX-1 服务器产生的深度学习训练工作负载, HA 对可剩余大约 70% 的性能余量。因此, AFF A800 系统可以支持更多 DGX-1 服务器。对于具有更高存储性能需求的更大规模部署, 则可以在 NetApp ONTAP 集群中增加更多 AFF A800 系统。ONTAP 9 支持一个集群使用最多 12 个 HA 对 (24 个节点), 而在此解决方案中验证的 FlexGroup 技术, 可以在一个卷中提供超过 20 PB 容量。我们在此次验证中使用的数据集相对较小。但是, ONTAP 9 可通过线性性能扩展将容量扩展到令人惊叹的程度, 因为每个 HA 对均可提供与本文档中验证的级别相当的性能。

对于更小规模的 DGX-1 集群, 更低价位的 AFF A220 或 AFF A300 系统便能提供充足的性能。由于 ONTAP 9 支持混合模式集群, 因此您可以从更小占用空间起步, 然后随着您的容量和性能需求的增长在集群中添加更多或更大的存储系统。

从网络角度来看, 按照验证结果, 此架构仅使用了每个 Nexus 3232C 交换机上 32 个可用端口中的 16 个端口。每个交换机可以使用额外存储访问端口支持多达八个 DGX-1 服务器, 因此无需增加网络便可大幅提升计算能力。对于更大实施规模, Cisco Nexus 7000 的每个交换机最多可支持 192 个线速 100GbE 端口。或者, 您也可以使用连接到中央主干交换机的多对 Nexus 3000 交换机实施分支-主干拓扑结构。

根据使用此人工智能训练工作负载进行的验证测试, 每个 DGX-1 均需要大约 2 GB/秒存储吞吐量。AFF A800 系统已证明在使用通过其他方式产生的类似工作负载时能够提供 25 GB/秒的吞吐量, 因而此架构中的每个 AFF A800 HA 对可支持九个或更多 DGX-1 服务器。

## 8 结论

DGX-1 服务器是功能极为强大的深度学习平台, 它可以利用同样强大的存储和网络基础架构来提供最大价值。通过将 NetApp AFF 系统与 Cisco Nexus 交换机相结合, 您几乎能够以所需的任何规模 (从一个与 AFF A220 系统配对的 DGX-1 服务器, 直到含 12 个节点的 AFF A800 集群上的 96 个 DGX-1 服务器) 实施经过验证的这种架构。AFF 与 NetApp ONTAP 的卓越云集成功能以及软件定义的功能相结合, 可为成功实施深度学习项目提供跨边缘、核心和云的完整数据管道。

## 致谢

我们要对来自 NVIDIA 令人尊敬的同行 Darrin Johnson、Tony Paikeday、Robert Sohigian 和 James Mauro 对这个经验证的 NetApp 架构所做的贡献表示衷心的感谢。没有 NetApp 关键团队成员 Robert Franz 和 Kesari Mishra 的支持和指导, 我们就无法完成此项研究。

我们要对以自己的真知灼见为本白皮书的研究带来巨大帮助的所有人士表示真诚的赞赏和谢意。

## 如何查找其他信息

如欲更详细地了解本文档所述的信息, 请参见下列资源:

- NVIDIA DGX-1 服务器
  - NVIDIA DGX-1 服务器  
<https://www.nvidia.com/en-us/data-center/dgx-1/>
  - NVIDIA Tesla V100 Tensor 核心 GPU  
<https://www.nvidia.com/en-us/data-center/tesla-v100/>
  - NVIDIA GPU Cloud  
<https://www.nvidia.com/en-us/gpu-cloud/>
- NetApp AFF 系统
  - AFF 产品规格  
<https://www.netapp.com/cn/media/ds-3582.pdf>

- NetApp 借助 AFF 展现闪存优势  
<https://www.netapp.com/us/media/ds-3733.pdf>
- ONTAP 9.x 文档  
<http://mysupport.netapp.com/documentation/productlibrary/index.html?productID=62286>
- NetApp FlexGroup 技术报告  
<https://www.netapp.com/cn/media/tr-4557.pdf>
- NetApp 互操作性表：
  - NetApp 互操作性表工具  
<http://support.netapp.com/matrix>
- Cisco Nexus 网络  
以下链接提供了有关 Cisco Nexus 3232C 系列交换机的详细信息：
  - Cisco Nexus 3232C 系列交换机  
<https://www.cisco.com/c/en/us/products/switches/nexus-3232c-switch/index.html>
  - Cisco Nexus 3232C 配置指南  
<https://www.cisco.com/c/en/us/support/switches/nexus-3000-series-switches/products-installation-and-configuration-guides-list.html>
  - Cisco Nexus 3232C 命令行参考  
<https://www.cisco.com/c/en/us/support/switches/nexus-3000-series-switches/products-command-reference-list.html>
- 机器学习框架：
  - TensorFlow:TensorFlow：适合所有人的开源机器学习框架  
<https://www.tensorflow.org/>
  - Horovod:Horovod：Uber 适用于 TensorFlow 的开源分布式深度学习框架  
<https://eng.uber.com/horovod/>
  - 在容器运行时生态系统中启用 GPU  
<https://devblogs.nvidia.com/gpu-containers-runtime/>
- 数据集与基准测试：
  - ImageNet  
<http://www.image-net.org/>
  - TensorFlow 基准测试  
<https://www.tensorflow.org/performance/benchmarks>

## 附录

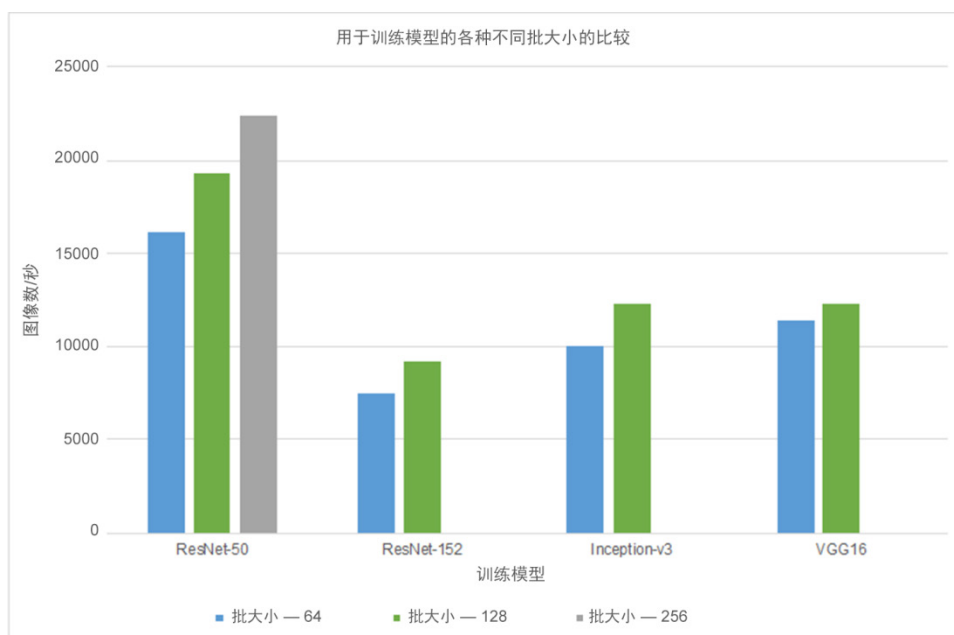
本节包含使用此架构执行的测试的更多结果。

### 每个模型使用不同批大小所达到的训练速度

图 16 显示了使用以下组件的不同训练模型使用各种不同批大小时的比较：

- GPU 数量：32 个（4 个 DGX-1 服务器）
- 核心：Tensor 核心
- 批大小：64、128 和 256（对于 ResNet-50）；64 和 128（对于其他模型）

图 16) 训练模型使用各种不同批大小时的比较。



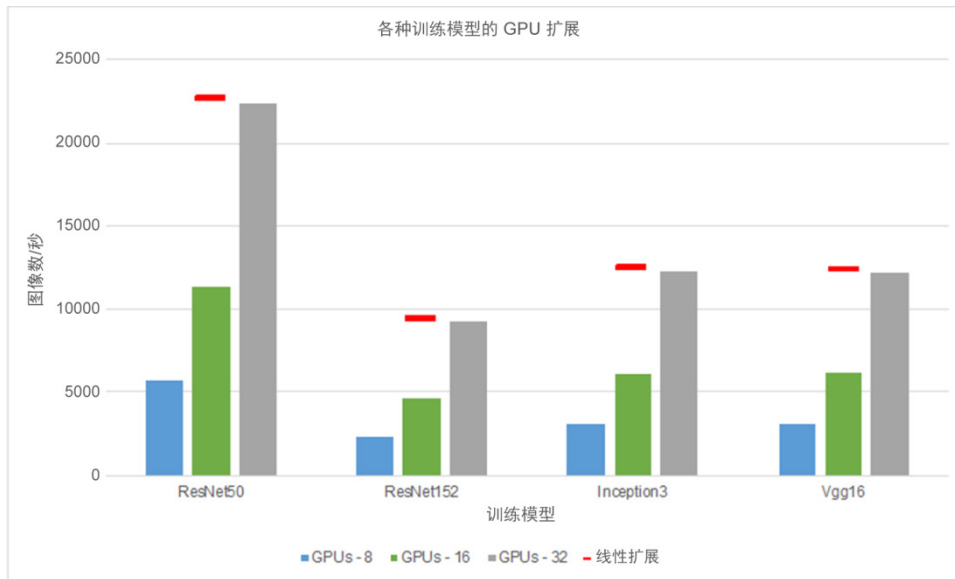
**结论：**当批大小提高到 256 或 128 时，训练吞吐量性能也随之提高。

### 每个模型的 GPU 扩展比较

图 17 显示了使用以下组件的不同训练模型的 GPU 扩展：

- GPU 数量：8 个（1 个 DGX-1 服务器）、16 个（2 个 DGX-1 服务器）和 32 个（4 个 DGX-1 服务器）
- 核心：Tensor 核心
- 批大小：ResNet-50 使用 256，其他模型使用 128

图 17) 各种训练模型的 GPU 扩展。



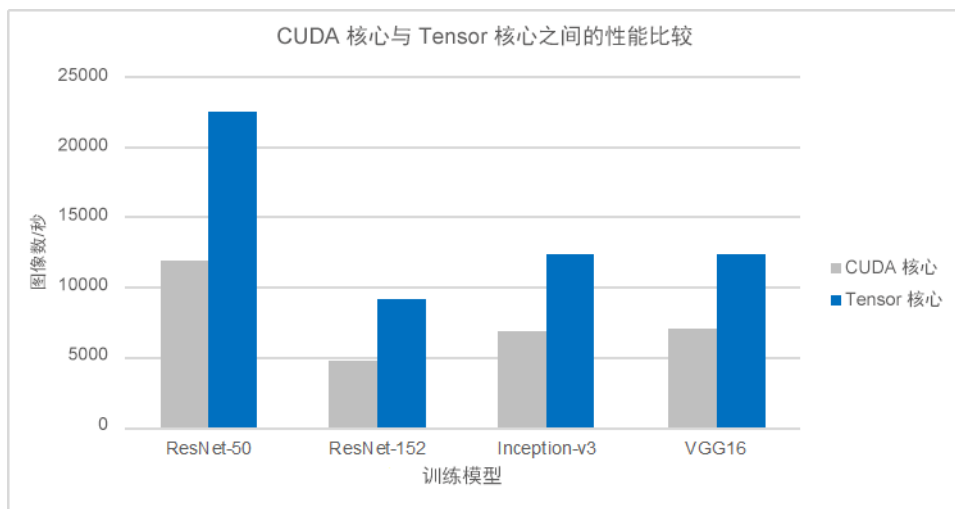
- 结论：所有训练模型的 GPU 扩展均为线性扩展。

## Tensor 核心和 CUDA 核心的比较

图 18 显示了使用以下组件的 CUDA 核心与 Tensor 核心之间的性能比较：

- GPU 数量：32 个（4 个 DGX-1 服务器）
- Cores:核心：Tensor 核心和 CUDA 核心
- 批大小：ResNet-50 使用 256，其他模型使用 128

图 18) CUDA 核心与 Tensor 核心之间的性能比较。



结论：Tensor 核心产生的性能优于 CUDA 核心产生的性能。

## 所有模型的 GPU 工作负载

图 19 至图 21 分别显示了使用以下组件的 ResNet-50、ResNet-152 和 Inception-v3 的 GPU 利用率和带宽：

- GPU 数量：32 个（4 个 DGX-1 服务器）
- 核心：Tensor 核心
- 批大小：ResNet-50 使用 256，其他模型使用 128

图 19) ResNet-50 的 GPU 利用率和存储带宽。

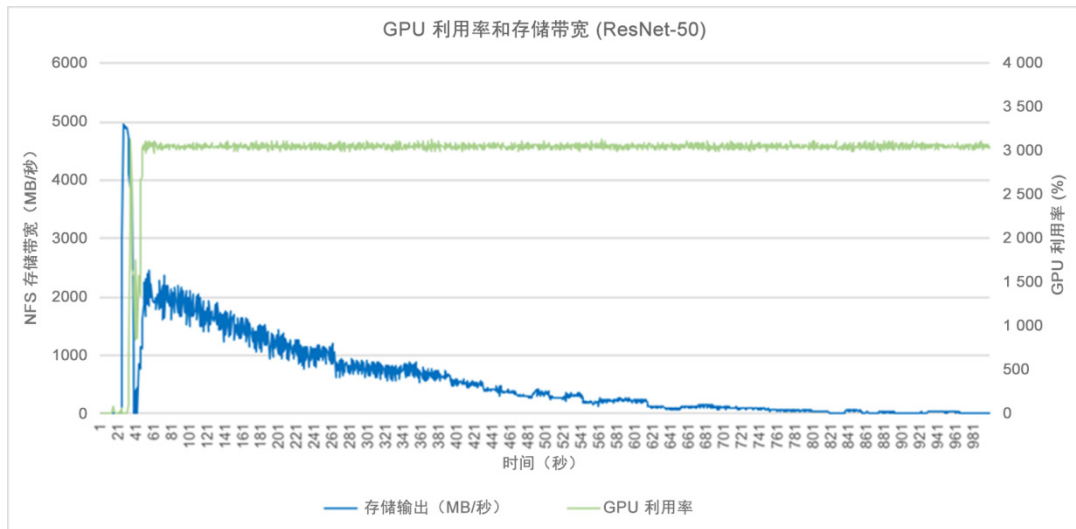


图 20 显示了 ResNet-152 的 GPU 利用率和带宽。

图 20) ResNet-152 的 GPU 利用率和存储带宽。

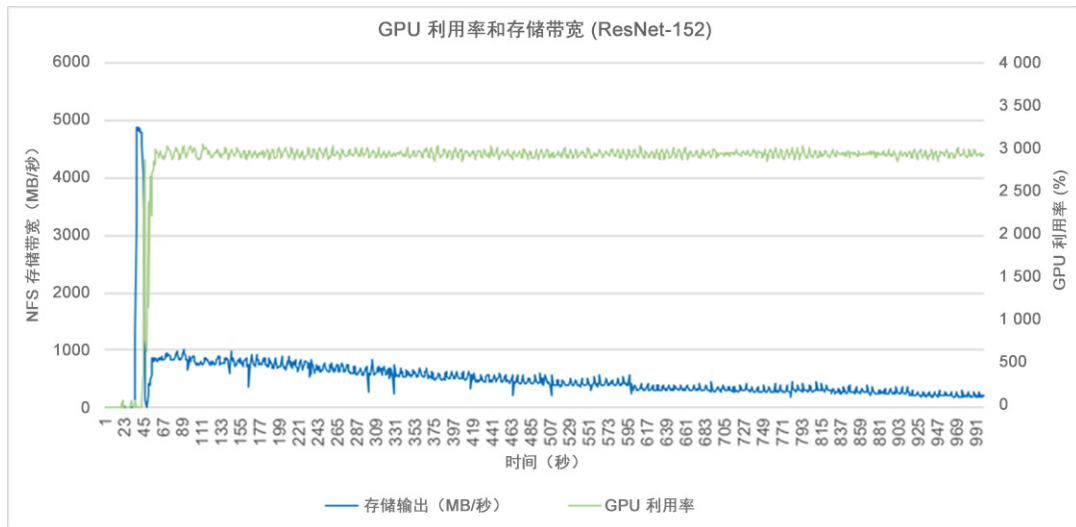
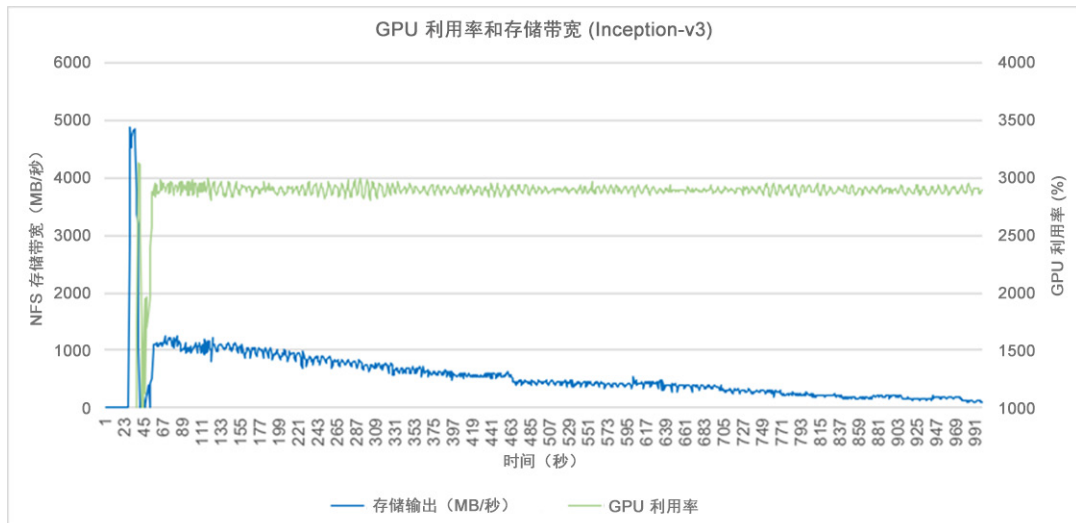


图 21 显示了 Inception-v3 的 GPU 利用率和带宽。

图 21) Inception-v3 的 GPU 利用率和存储带宽。





要验证您的特定环境是否支持本文档所述的确切产品和功能版本，请参见 NetApp 支持站点上的[互操作性表工具 \(Interoperability Matrix Tool, IMT\)](#)。NetApp IMT 中定义的产品组件和版本可用于构建 NetApp 所支持的配置。具体的配置结果取决于每个客户如何依照所发布规格进行安装。

## 版权信息

版权所有 © 1994 - 2018 NetApp, Inc.。保留所有权利。未经版权所有者事先书面许可，本文档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国

本文档中所含数据与商用项目（按照 FAR 2.101 中的定义）相关，属于 NetApp, Inc. 的专有信息。美国政府对这些数据拥有非排他性、不可转让、无转授权、全球性、受限废除的许可，但仅限于依据美国政府合同提供的数据以及为此类数据提供支持所需。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b) 条款中明确的权利。

## 商标信息

NetApp、NetApp 标识和 <http://www.netapp.com/TM> 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。