

백서

자율주행차 소프트웨어 개발 기업

퓨어스토리지와 NVIDIA의 딥러닝 분야
고객 레퍼런스 아키텍처

개요

인공지능(AI)은 다양한 산업에서 새로운 혁신의 물결을 만들어 내고 있습니다. 머신러닝의 가장 흥미로운 활용 사례는 자율주행차입니다. 불과 몇 년 전만 하더라도, 자율주행차와 관련된 아이디어는 먼 미래의 이야기로 인식되었습니다. 그러나, 고급 훈련 알고리즘의 부상과 컴퓨터 및 빅데이터 분야의 발전 덕분에 자동차 업계는 공상과학 소설과 같은 자율주행 개념을 현실화할 수 있게 되었습니다. 자율주행차의 실현을 위해서는 방대한 양의 데이터가 트레이닝 된 알고리즘 및 컴퓨팅 시스템에 적용되어야 하기 때문에 현대적인 분석을 위해 특별히 설계된 스토리지 솔루션이 필요합니다.

“교통 사고의 94%는 운전자의 실수로 인해 발생한다. 자율주행차의 도입은 이러한 사고를 획기적으로 줄여 줄 수 있는 엄청난 혁신을 의미한다.”

—
AUDI

NVIDIA® 와 퓨어스토리지는 기업들이 프라이빗 클라우드 환경에서 딥러닝 및 인공지능의 강력한 성능을 활용할 수 있도록 지원하기 위해 협력하고 있습니다. 본 자료는 NVIDIA DGX-1™과 퓨어스토리지의 플래시블레이드(FlashBlade™)로 구축한 인공지능 인프라를 기반으로 향후 5년 안에 완전 자율주행 솔루션을 상용화할 예정인 고객의 사례를 다룬 백서입니다.

자동차 시장에서 자율주행을 둘러싼 경쟁이 점차 치열해지고 있으며, 자율주행 상용화를 위해 개발을 가속화하는 일은 매우 중요해지고 있습니다. 원천에 위치한 자동차 컨설팅 기업 Beryll Strategy Advisors의 총괄 잰 버가드(Jan Burgard)는 다음과 같이 말합니다.

“자율주행차 개발에 있어서는 선두 주자가 되는 것이 정말로 대단히 중요하다. 자율주행 기술이 자동차 제조업체의 핵심 비즈니스를 크게 변화시키는 중요 요소가 될 것이 명백하기 때문이다. 선두 주자가 된다는 것은 이러한 변화를 평가, 관리하는 동시에 미래 개발 혁신을 선도할 수 있는 전략을 갖추는 것을 의미한다.”

자율주행은 다수의 엔지니어링 역량을 갖춘 인재, 막대한 컴퓨팅 성능, 그리고 지속적으로 확장 가능한 대규모의 현대적인 데이터 플랫폼을 요구하는 어려운 도전 과제입니다. 본 자료에서 다루는 고객 사례에서 고객은 자사의 미션 달성을 돕는 인프라를 가장 잘 구축하기 위해 다양한 옵션들을 고려 및 평가한 끝에 NVIDIA DGX-1™과 퓨어스토리지의 플래시블레이드(FlashBlade™)를 선택했습니다. 본 백서에서는 고객이 두 가지 솔루션 도입을 최종 결정하게 된 기술적인 요인들에 대해 집중적으로 살펴 보도록 하겠습니다.

고객 요구사항

이 고객은 개발 팀들이 전세계에 분산되어 있으나, 한 곳에 머신러닝을 위한 통합 시스템을 구축하여 각 팀의 개발 단계에서 머신러닝을 사용하고 있습니다. 본 통합 시스템의 구현에 NVIDIA DGX-1과 퓨어스토리지 플래시블레이드(FlashBlade)를 선택했습니다. 그 이유는 레거시 솔루션이 수십 개의 랙을 사용하는 반면 이 두 가지 솔루션은 단 몇 개의 랙 유닛에 통합된 고밀도 폼팩터에 컴퓨트 및 스토리지를 제공하기 때문입니다. 본 시스템에서는 데이터센터 상면 공간을 최적화할 뿐 아니라, 전력 및 냉각 비용을 최소화하는 것 또한 핵심 과제였습니다. 솔루션의 이 같은 장점은 매우 중요한 부분이었습니다.

주요 데이터 세트에는 평균 1MB에서 2.5MB 크기의 수 천만 개의 JPG 이미지들이 포함되어 있습니다. 이 이미지들은 해당 시스템에서 대규모 모델의 트레이닝 뿐만 아니라 프로세싱 및 탐색 작업에 활용됩니다. (보다 자세한 내용은 '아키텍처 개요' 부분에서 확인할 수 있습니다.)

자율주행 소프트웨어 개발을 위한 통합 시뮬레이션 테스트 과정은 개별 GPU에 학습을 위하여 대량의 데이터 세트를 무작위로 지속적으로 빠르게 제공해야 합니다. 대규모의 데이터 세트에서 트레이닝용 이미지를 랜덤 하게 읽게 되면 캐시 미스가 높아져서 기존 스토리지가 가지는 캐시 메커니즘을 수행할 필요가 없습니다. 따라서 스토리지 어플라이언스에서 데이터를 빠르게 읽는 것이 훨씬 더 효율적이었습니다. 플래시블레이드(FlashBlade) 기반으로 구축하기 이전에, 이 고객은 화이트박스 서버에서 구현한 JBOB 및 경쟁업체의 스케일-아웃 스토리지 등 다양한 스토리지 솔루션을 테스트해 봤습니다. 그러나 어떠한 다른 스토리지도 GPU 성능을 최대한 사용하게 하지 못하였습니다.

이와 같은 AI 슈퍼 컴퓨터를 위한 또 다른 핵심 요구사항은 간단함이었습니다. 데이터 과학자들은 인프라 관리가 아닌 모델 튜닝 및 트레이닝에 시간을 투자하길 원합니다. DGX-1은 디버닝을 위해 사전에 최적화되어 구성된 프레임워크 및 툴을 제공하기 때문에 디버닝 워크플로우를 위한 별도의 환경 튜닝이 거의 필요하지 않습니다. 이와 마찬가지로, 플래시블레이드(FlashBlade) 또한 모든 트레이닝 과정에서 관리자의 별도의 세팅없이 업무환경 전반의 모든 훈련 노드에 데이터를 균일하게 전송하여 업계 최고 수준의 간단한 관리를 제공 합니다. 이는 다음과 같은 두 가지의 관점에서 효율성을 향상시켜 줍니다. 첫째, 훈련 구동을 위해 데이터를 지속적으로 복사하거나, 둘째, 극대화된 성능을 확보하기 위해서 하드웨어를 지속적으로 튜닝할 필요가 없어집니다. 이 두가지 관점의 효율성은 데이터 과학자들이 데이터 관리를 보다 쉽게 합니다.

아키텍처 개요

본 아키텍처 설계의 두 가지 주요 고려사항은 다음과 같습니다. 첫째로, 프로젝트의 반복성을 지원할 수 있는 역동적이고 민첩한 인프라를 구축하는 것이며, 둘째로는 이 고객의 AI 기반 솔루션 및 역량이 발전함에 따라 이에 맞는 컴퓨트 및 스토리지 투자를 준비하는 것입니다. 고객은 기존 다수의 디스크를 사용하는 하둡 스타일의 정적 솔루션이 더 이상의 대안이 될 수 없음을 인지하고, 디버닝 기반으로 개발 및 테스트하는 환경에서 유연하게 스케일 업 또는 다운될 수 있는 솔루션이 필요로 하였습니다.

다음의 아키텍처 다이어그램(그림 1)은 고객이 디버닝 인프라의 초기 구축에 구현한 아키텍처입니다. NVIDIA DGX-1 노드의 클러스터링에 대한 보다 자세한 정보는 [여기](#)에서 확인할 수 있습니다.

이 자율주행 소프트웨어 기업의 데이터 과학자들은 클러스터링 된 DGX-1 인스턴스뿐 아니라, 완전한 훈련 시뮬레이션을 수행하기 이전의 반복적 개발 작업에 NVIDIA Tesla P100가 장착된 다수의 GPU 서버와 CPU 서버를 활용하고 있습니다.

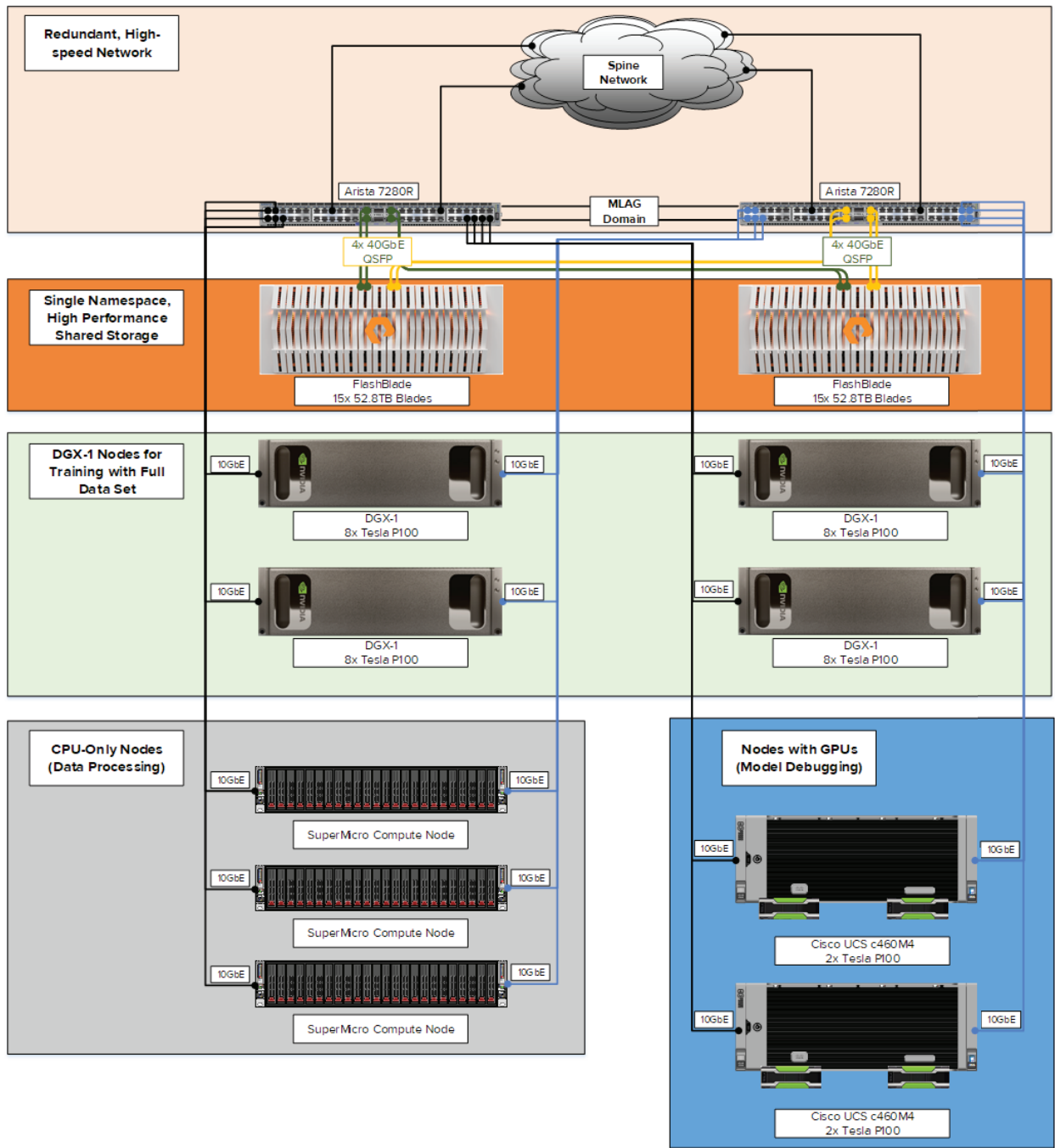


그림 1. 자율주행차 소프트웨어 기업의 아키텍처 다이어그램

그림1의 각 섹션에 대한 보다 자세한 사항은 다음과 같습니다.

네트워크: 이 다이어그램의 상단은 향후 성장을 위해 충분한 대역폭을 제공하는 이중화된 고속 10/40/100Gb 이더넷 네트워크 스위치 한 쌍을 보여줍니다. Arista 스위치들은 MLAG 도메인을 통해 페어링되고 모든 컴퓨터와 스토리지는 각 스위치에 연결되어, 단일 포트 또는 스위치에 문제가 발생하는 경우를 위한 이중화와 적절한 로드 밸런싱을 제공합니다. AI 훈련에 사용되는 다수의 컴퓨터 노드들은 모두 플래시블레이드(FlashBlade)에 저장된 훈련용 데이터 세트에 액세스할 수 있습니다.

플래시블레이드(FlashBlade): 플래시블레이드(FlashBlade)는 기존의 스토리지와는 차원이 다른 스토리지 시스템입니다. 소프트웨어부터 하드웨어에 이르는 방대한 병렬 아키텍처를 갖춘 플래시블레이드(FlashBlade)는 AI 데이터 파이프라인의 비정형 데이터 워크로드에 전혀 없는 성능을 제공합니다.

앞서 언급했던 것처럼, 이 자율주행 소프트웨어 기업의 데이터 과학자들은 클러스터링된 DGX-1 인스턴스만 사용하는 것이 아니라 완전한 트레이닝 시뮬레이션을 생성하기 이전 반복적인 개발 작업에 다양한 유형의 노드를 활용합니다. 데이터 인제스트에서부터 CPU 전용 서버를 통한 데이터 처리, 경량 GPU 서버를 사용한 모델 디버깅 및 DGX-1 서버를 통한 실제 훈련 구동에 이르기까지, 플래시블레이드(FlashBlade)는 고성능 데이터 허브를 제공합니다. 이로써 데이터 과학자들은 가용 노드를 통해 데이터 세트를 계속 이동할 필요 없이, 동일한 데이터의 복사본을 여러 개 생성할 수 있습니다.

이보다 더욱 중요한 점은 트레이닝 구동 중에 SSD 또는 컴포넌트가 고장이 나더라도 데이터 또는 작업이 손실되지 않는다는 사실입니다. 또한, 플래시블레이드(FlashBlade)는 훈련용 데이터 세트에 더 많은 데이터가 추가됨에 따라, 일관적인 선형적 대역폭 성능을 제공합니다. 향후 데이터 세트의 규모와 훈련 알고리즘의 규모 및 정교함이 증가함에 따라, 스토리지 용량 및 성능 요구사항을 예측하고 정확한 예산을 수립할 수 있도록 합니다.

NVIDIA DGX-1 클러스터: 이 시스템의 목표는 훈련을 실행하는 동안에는 트레이닝용 데이터를 활용해 컴퓨팅 시스템을 끊임 없이 구동하는 것입니다. DGX-1이 이러한 방대한 데이터 세트를 인제스트 하고 훈련시키는 엔진의 역할을 한다면, 플래시블레이드(FlashBlade)는 연료를 공급하는 역할을 합니다. 플래시블레이드(FlashBlade)와 마찬가지로, DGX-1 노드를 기존 클러스터에 추가하면 전체 처리 성능이 획기적으로 향상됩니다. 이를 통해 이 고객은 물론 다른 고객들 또한 우선 자사의 요구사항에 맞는 규모로 시스템을 구축한 후, 컴포넌트가 포화상태에 도달한다면 노드를 추가하여 간단하게 성능을 확장시킬 수 있습니다. 이 고객은 4개의 DGX-1 시스템을 현재 사용하고 있으며, 추후 컴퓨터 용량을 확장해야 할 경우 플래시블레이드(FlashBlade) 또한 성능 저하 또는 시스템 중단 없이 아주 간단하게 확장 가능합니다.

GPU 가속 노드 및 CPU 전용 노드: 일반적으로 엔지니어링 프로젝트에서 종합적인 결과를 얻기 위하여 많은 시간을 하부 컴포넌트들을 작업하는 데 할애하고 있습니다. 데이터 과학자들도 이와 비슷한 과정을 거쳐야 하므로 실제 DGX-1 클러스터에 접속하여 사용하는 시간보다 다음의 세가지 작업에 가장 많은 시간을 할애 합니다.

- 트레이닝용 데이터의 수집, 분석, 정제, 필터링 및 처리 및 모델 훈련에 적합한 형태로의 변환
- 트레이닝용 데이터 소규모 하부 세트를 통한 모델 테스트 및 디버깅
- 전체 트레이닝용 데이터 세트를 통한 모델 훈련

처음의 두 가지 작업은 주로 상용 CPU 또는 GPU 서버에서 수행됩니다. 기존은 위 세가지 작업이 독립적인 저장 공간에서 진행되어 다음 작업을 위해서 데이터 이동이 발생하였고, 데이터 이동에 필요한 관리와 시간이 필요하였습니다. 그러나 플래시블레이드(FlashBlade)는 독립적인 저장소를 범용 데이터 허브로 제공하여 이 세 가지 작업들 간의 데이터 이동을 최소화함으로써 생산성을 더욱 향상시킵니다.

벤치마크 테스트 설정

퓨어스토리지는 업계 표준 벤치마크에 기반한 결과를 제공하기 위해 DGX-1과 플래시블레이드(FlashBlade)의 조합을 내부적으로 벤치마킹 하는데 많은 노력을 기울였습니다. 물론 잠재 고객들에게는 DGX-1 및 플래시블레이드(FlashBlade)의 최적의 조합을 판단하게 하기 위해 자체 데이터와 모델을 활용해 테스트를 수행할 것을 권장합니다. 이 결과들은 전체 벤치마크 테스트 수행의 하부 세트에 불과하며, 보다 자세한 정보를 원하는 고객은 퓨어스토리지 담당자에게 연락해 주시기 바랍니다.

내부 테스트에서는 P100 GPU 8개를 탑재한 DGX-1 서버 한 대(Pascal 아키텍처 기반 시스템)에 CUDA 코어 28,672개, 호스트 CPU 2개(총 80개 코어) 및 시스템 메모리 512GB를 조합했습니다. 참고로, V100 기반의 DGX-1 버전을 사용하면 보다 향상된 결과를 얻을 수 있을 것으로 예상됩니다. 트레이닝용 데이터는 15개 블레이드로 구성된 플래시블레이드(FlashBlade)에 저장되었으며, 모든 테스트는 NVIDIA가 공급한 컨테이너(nvcr.io/nvidia/tensorflow v17.07)에서 수행되었습니다.

훈련용 데이터의 입력을 위해서는 ImageNet 2012 데이터 세트를 사용했습니다. 이 데이터 세트는 [딥러닝 연구에서 가장 보편적으로 사용되고 있는 데이터 세트](#)입니다. 150KB 크기의 JPG 이미지들을 보다 큰 크기의 파일(각 최대 135MB)로 결합하여 입력 데이터로 사용했으며, 파일 시스템 캐싱(fsc) 옵션은 사용하지 않았습니다. 또한 TensorFlow 성능 최적화를 위한 [베스트 프랙티스](#)에 따라, 초 당 처리되는 이미지의 개수가 안정적으로 유지될 때까지 각 모델을 트레이닝시켰습니다.

DGX-1에 있는 두 개의 10Gbps 링크를 모두 활용하기 위해, 다수의 TCP를 연결해 입력 데이터에 액세스할 필요가 있었습니다. 플래시블레이드(FlashBlade) 파일 시스템을 두 개의 서로 다른 데이터 VIP에 마운트하고 훈련 과정에서 두 마운트 지점의 파일 읽기 액세스를 하나로 병합했습니다.

벤치마크 결과

다음 그래프는 TensorFlow 데이터 세트를 사용하여 4가지 모델을 벤치마크한 결과로, 각 모델 별로 두 가지 유형의 결과를 보여줍니다. 첫번째 테스트 즉 Synthetic 테스트는 시스템 RAM에서 데이터가 무작위로 생성되어 입력되는 조건 하에서 GPU의 쓰루풋을 테스트하도록 설계되었습니다. 이러한 설정에서는 I/O를 플래시블레이드(FlashBlade)로 전송하거나, 실제 ImageNet 데이터 세트를 사용할 필요가 없으며, 종합 테스트 결과는 달성할 수 있는 최고의 훈련 성능의 절대적인 수치를 나타냅니다. 두번째 테스트는 플래시블레이드(FlashBlade)에 저장된 ImageNet 데이터 세트를 활용해 테스트하여 얻은 결과입니다. 이 첫번째 두번째 테스트 결과를 비교해 보았습니다.

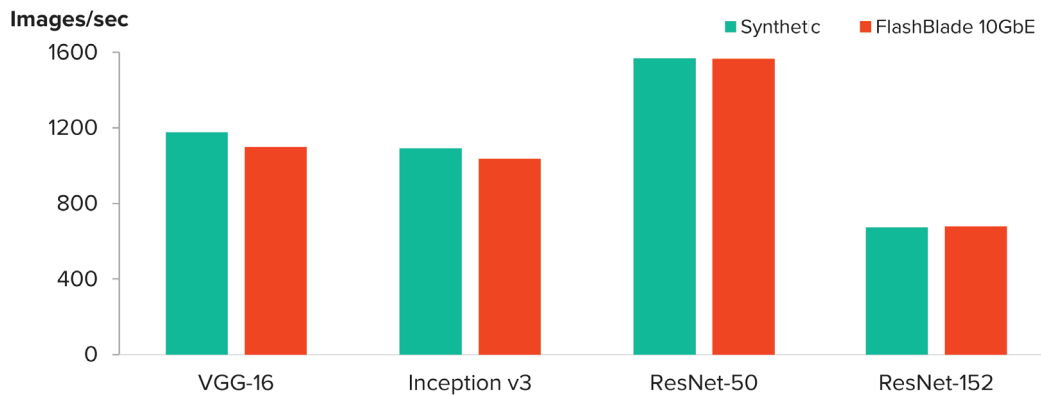


그림 2. 벤치마킹 결과

시스템 RAM으로부터 데이터가 직접 읽기 되는 종합 데이터와의 비교 테스트 결과, 플래시블레이드(FlashBlade)는 이와 유사한 딥러닝 훈련 성능을 제공하는 것으로 나타났습니다. 플래시블레이드(FlashBlade) 설정에서 데이터는 네트워크 패브릭에 있는 외부 스토리지에서 PCIe 익스프레스 기반으로 상호 연결되고, 네트워크 인터페이스 카드(NIC)를 통해 읽어와야 하는 것을 감안할 때, 이는 매우 놀라운 결과라고 할 수 있습니다. 이러한 테스트 결과는 플래시블레이드(FlashBlade)가 딥러닝 훈련에 최고의 성능을 제공한다는 사실을 보여줍니다.

ImageNet은 광범위하게 사용되는 벤치마킹 툴로서 다양한 프레임워크와 솔루션들을 비교해 볼 수 있는 방법을 제공합니다. 그러나 고객의 실질적인 요구사항이 반영되지 않는 경우가 많습니다. 예를 들어, 이 고객은 ImageNet 데이터 세트에서 150KB 보다 훨씬 큰 크기의 파일들을 사용합니다. 이에 파일 크기에 대한 광범위한 요구사항에 따른 성능을 이해하는 것이 중요합니다. 공유 스토리지로 확보될 수 있는 성능은 입력 파일의 크기에 따라 달라집니다.

- 작은 크기의 파일(50KB)의 경우, 7개 블레이드로 구성된 플래시블레이드(FlashBlade) 시스템은 무작위 데이터에 대해 약 5GB/초의 읽기 쓰루풋을 제공할 수 있으며, 15개 블레이드로 구성된 시스템이 제공할 수 있는 읽기 쓰루풋은 약 10GB/초 정도로 제공됩니다.
- 보다 큰 크기의 파일(1MB 이상)의 경우, 읽기 성능은 블레이드 당 1GB/초를 약간 상회하여, 15개 블레이드 시스템의 경우 15GB/초의 성능을 발휘합니다.

향후, V100처럼 더 강력한 성능을 제공하는 GPU가 출시되면 P100의 훈련 속도가 두 배 이상 빨라지고, 스토리지 시스템 성능에 대한 수요 또한 증가할 것입니다. 또한, 더욱 강력한 성능을 제공하는 GPU는 더 심층적인 신경망을 가능하게 하여 입력 배치(batch)당 계산량을 증가시킵니다. 보다 심층적인 모델은 성능에 대한 수요 증가량을 일부 상쇄시킬 것으로 예상됩니다.

결론

AI와 딥러닝 분야에 대해 공통적으로 기대되는 것은 변화와 진보를 통한 끊임 없는 한계 돌파로, 자율주행차와 같은 분야에 흥미롭고 새로운 활용 사례와 발전을 촉진할 것이라는 점입니다. 지속적으로 진화하는 딥러닝을 지원하려면, AI 데이터 파이프라인을 위한 기반 인프라는 유연하고 선형적 확장이 가능해야 하며, 간단하게 관리될 수 있어야 합니다. 업계의 다양한 옵션들을 검토한 후, 이 고객은 업계 최고의 컴퓨트 및 스토리지 플랫폼이 결합된 NVIDIA와 퓨어스토리지의 솔루션을 선택했습니다. 이러한 두 가지의 솔루션은 이 기업이 자율주행차 혁명을 선도하고자 하는 목표를 실현할 수 있게 도와줄 것입니다.

© 2017 Pure Storage, Inc. All rights reserved.

Pure Storage, FlashStack, 및 Pure Storage 로고는 미국 또는 그 외 국가에 위치한 자회사의 상표 또는 등록상표입니다. Nvidia는 Nvidia, Inc.의 등록상표입니다. 기타 기업, 제품 또는 서비스 이름은 해당 기업의 등록상표입니다.

이 문서에 설명된 퓨어스토리지 제품들은 제품의 사용, 복사, 배포 및 역컴파일/역엔지니어링을 제한하는 라이선스 계약 하에 배포됩니다. 이 문서에 설명된 퓨어스토리지 제품들은 라이선스 계약의 조건에 따라서만 사용될 수 있습니다. 이 문서의 어떠한 부분도 퓨어스토리지의 사전 서면 허가 없이 어떠한 형식이나 방법으로도 복제될 수 없습니다. 퓨어스토리지는 이 문서에 포함된 퓨어스토리지의 제품 및/또는 프로그램을 사전 통지 없이 언제든지 임의대로 개선 및/또는 변경할 수 있습니다.

이 문서는 “있는 그대로” 제공되며, 퓨어스토리지는 법적으로 허용된 범위 내에서 상품성, 특수 목적을 위한 적합성, 또는 비침해성에 대한 보증은 물론 그 어떠한 명시적, 묵시적, 서면, 구술 또는 법적 보증을 부인합니다. 퓨어스토리지는 이 문서의 이용, 공급 또는 성과와 관련하여 발생하는 모든 우발적 또는 결과적 손해에 대해 어떠한 경우에도 책임을 지지 않습니다. 이 문서의 정보는 예고 없이 변경될 수 있습니다.

ps_wp7p_dgx1-flashblade-anonymous_ltr_01