B2B 및 B2C 비즈니스를 위한 첨단 AI 및 딥 러닝 기술 공급업체 지프(ZIFF)는 퓨어스토리지의 플래시블레이드(FlashBlade™)를 사용해 가장 까다로운 AI 워크로드를 지원하는 핵심 데이터 플랫폼을 구축함으로써 전례 없는 속도로 데이터를 처리할 수 있게 되었습니다.

ZIFF

비즈니스 혁신

지프는 전례 없던 수준의 AI와 딥 러닝 서비스를 제공하여, 고객들이 수개월의 시간과 수십만 달러의 비용을 절감하고 글로벌 기업들도 해결할 수 없었던 문제들에 대한 답을 찾을 수 있도록 지원해줍니다.

지역

북미

산업

IT

" 지프는 AI를 혁신적으로 사용해 CEO들이 다시 한번 자신들의 비즈니스에 열정을 가질 수 있도록 만들고자 한다."

데이빗 곤잘레스(David Gonzalez), **CEO**

플래시블레이드(FLASHBLADE)를 통한 AI 분석의 장애물 제거

혁신적인 CEO들 사이에 "매너리즘에 빠진 비즈니스 리더들은 역사를 만들 수 없다"라는 격언이 있습니다. 기존의 틀을 깨는 사고를 하고 고정된 비즈니스 관행의 프레임을 벗어나 새로운 인사이트를 확보하는 이들이 기존 업계를 뒤흔들거나 새로운 업계를 만듭니다. 성공을 위해 새로운 방법을 모색하는 CEO들이 AI와 딥 러닝으로 눈을 돌리고 있는 이유도 바로 이 때문입니다. 그렇지만, 글로벌 기업들을 포함한 대부분의 조직은 AI와 딥 러닝 기법을 기업의 비즈니스에 적용하는데 필요한 인적 및 물적 자원을 갖추고 있지 않습니다. 기업들이 이러한 장애물을 극복할 수 있도록 지원하는 것이 지프(ZIFF Inc.)의 목표입니다.

지프의 공동 창립자 겸 CEO인 데이빗 곤잘레스(David Gonzalez)는 "지프는 대부분의 기업이 사용자 경험을 향상시키는데 필요한 AI나 머신 러닝 분석 역량을 보유하고 있지 않다는 가정에서 출발했다"고 말했습니다. 또 다른 공동 창립자이자 최고 AI 책임자인 벤 테일러(Ben Taylor)는 "여기에는 여러 가지 이유가 있다. 선택해야 할 프레임워크가 많고 과정도 매우 복잡하며, 대기업들조차도 필요한 인재를 모두 보유한 경우는 흔치 않다. 기업들은 자체적으로 이러한 작업을 수행하길 바라지만 인력이나 경험이 부족한 실정이다. Fortune 500 기업 중에도 2년 이상 프로젝트를 진행하고 있지만, 아직 결실을 맺지 못한 기업들이 있다"고 덧붙였습니다.

지프가 CEO들에 제안하는 가치는, 내부 분석을 위한 인력 이외 기타 자원이 부족한 기업들을 위해 AI 기술이 핵심 상품과 서비스에 향상에 도움이 될 수 있도록 합니다. 지프는 전문지식을 동원해, 총 3단계에 걸친 AI 프로젝트를 통하여 기업들을 지원합니다. 3단계는 다음과 같습니다. 첫째, 정확성을 제공할 수 있을 만한 분량의 데이터세트를 수집하고, 둘째, 수집된 데이터를 분석한 후, 마지막으로 분석 결과를 고객들이 해결하고자 하는 실질적 문제에 적용하는 것입니다.

일례로, 한 미국 자동차 보험회사의 CEO는 자동화된 프로세스를 통하여 자동차 손해 평가 시간과 비용을 줄이길 원했습니다. 이를 위해서는 4백만 장이 넘는 파손된 자동차 사진을 처리할 필요가 있었습니다. 보험사 직원들이 수년 동안 수집된 이 사진들을 식별해 라벨을 붙였습니다. 라벨을 붙인 데이터를 다시 2개월 동안 검토한 결과, 보험사의 데이터 과학자들은 이렇게 수작업으로 이루어진 라벨링은 AI 프로젝트의 기반이 될 수 있을 정도의 정확성을 갖추지 못했다는 결론을 내렸습니다.

이 고객은 지프의 데이터 큐레이션 소프트웨어를 이용해 4백만 장의 사진에 대한 라벨을 다시 생성했습니다. 결과물은 수작업 라벨보다 30% 더 정확했으며 작업 시간은 24시간도 채 걸리지 않았습니다. 새로운 알고리즘은 이 보험사 고객의 보험 청구 평가 절차에 적용되어, 기존 모델 대비 일일 3배에 달하는 견적을 생성할 수 있게 되었습니다.

고객명:

ZIFF
ziff.ai

활용 사례:

- 플래시블레이드(FlashBlade) – 인공 지능, 딥 러닝
- 데이터베이스 – Postgres®

도전 과제:

- 데이터 스토리지 I/O 내 지속적인 병목 발생
- 페타바이트급 데이터 파일이 스토리지 인프라스트럭처에 과도한 부담을 주는 문제
- 고객들의 보다 복잡한 질문에 대한 답의 필요성

IT 혁신:

- I/O에 대한 우려 사항 제거
- 지속적으로 높은 활용 수준을 유지하며 고가치의 데이터 과학자들의 생산성 향상
- 질문에 대한 답과 실행 가능한 인사이트 확보에 필요한 시간 및 비용 절감

> “ 플래시블레이드(FlashBlade)를 설치한 후 처리할 수 있는 워크로드가 3배 늘었다.”
>
> 벤 테일러(Ben Taylor), 최고 AI 책임자

병목현상으로 인한 문제

현재, 지프는 큐레이션 소프트웨어를 이미지, 비디오 및 오디오로 구성된 데이터세트에 적용하고 있습니다. 각 데이터세트는 테라바이트급으로, 페타바이트 규모가 되는 경우도 있습니다. 데이터의 효율적인 처리는 지프의 창립자들이 지프를 설립하기 훨씬 전부터 풀어야 할 숙제였습니다.

수석 데이터 과학자를 지낸 테일러는 다수의 GPU(그래픽처리장치), 하드 디스크, 솔리드 스테이트 스토리지 장치를 사용해 맞춤형 시스템을 구축했습니다. 그는 “흥미로운 딥 러닝 프로젝트와 훌륭한 연구자들을 보유하고 있었지만 언제나 스토리지 용량이 부족했다”며 “자신의 프로젝트를 위한 공간을 마련하기 위해 시스템에서 다른 연구자의 프로젝트를 삭제해 달라는 요청도 생기기도 했다. 이 모든 갈등을 조율해야 했고, 이 때문에 데이터를 시스템 안팎으로 이동하는 작업을 하며 작게는 수 시간, 길게는 며칠을 소비해야만 했다. 전혀 가치가 없는 이런 일로 연구자들은 스트레스를 받았다”고 말했습니다.

테일러는 “NFS 서버가 1대 있었는데 NFS 마운트가 항상 다운되는 등 온갖 문제가 발생했다”고 덧붙였습니다.

곤잘레스는 “데이터 I/O에 항상 신경을 곤두세우고 있어야 했으며 데이터를 신속하게 이동할 수 없어서 가장 중요한 리소스인 GPU와 데이터 과학자들의 능력을 활용할 수 없었다”고 회고했습니다.

플래시블레이드(FLASHBLADE), AI를 위한 현대적인 데이터 플랫폼

IT 인프라스트럭처를 구축하기 위해 여러 가지 스토리지 옵션을 검토하던 테일러는 지프의 고유한 니즈에 적합한 솔루션을 발견했습니다. 바로 퓨어스토리지의 플래시블레이드(FlashBlade)였습니다. 플래시블레이드(FlashBlade)는 데이터 분석, 인공 지능, 딥 러닝과 같이 방대한 병렬 문제를 처리하도록 특수 설계된 최초의 올플래시 데이터 플랫폼입니다.

최고의 I/O, 성능, 거의 무한대의 확장성, 운영 비용 절감을 지원하는 작은 설치 면적, 간편한 관리 등, 플래시블레이드(FlashBlade)는 현대적인 데이터 플랫폼에 대한 지프의 모든 기준을 충족했습니다. 지프는 플래시블레이드(FlashBlade)에 8TB 블레이드 7개를 설치해 70TB 이상의 유효 용량을 확보했습니다.

플래시블레이드(FlashBlade)는 지프의 비즈니스 운영에 유의미한 영향은 미쳤습니다.

곤잘레스는 “플래시블레이드(FlashBlade)가 I/O 이슈를 제거해주었다”며 “필요한 성능 그 이상을 체험할 수 있었고, 자동차 보험사 고객을 위해 24시간 만에 4백만 개의 이미지를 인덱싱할 수 있었다. 플래시블레이드(FlashBlade)가 없었다면 그렇게 짧은 시간에 완료한다는 것은 불가능했을 것이다”고 말했습니다.

처리 시간을 획기적으로 단축함으로써, 지프는 더 많은 고객 프로젝트를 맡을 수 있게 되었을 뿐 아니라 더 복잡한 프로젝트도 문제없이 처리할 수 있게 되었습니다. “플래시블레이드(FlashBlade)를 설치한 후 처리할 수 있는 워크로드가 3배 늘었다”고 테일러는 말했습니다. 그는 “이제 더 이상 클라우드나 SAN에서 데이터세트를 정렬할 필요가 없다. 플래시블레이드(FlashBlade)를 사용하면 수백만 장의 이미지를 한 시간 내에 로드할 수 있다. 전에는 훨씬 더 오래 걸리거나 아예 불가능했던 작업이다”라고 설명했습니다. 플래시블레이드(FlashBlade)는 시간을 절약해주고 생산성을 향상해주었습니다.

플래시블레이드(FlashBlade)는 지프의 “딥 인덱싱(deep indexing)”에도 큰 혜택을 주었습니다. 딥 인덱싱은 딥 러닝 기법을 이용해 짧은 시간 동안 정확하게 많은 양의 이미지를 분류 편집하는 것을 말합니다. “플래시블레이드(FlashBlade)는 엄청난 양의 데이터세트를 신속하게 처리할 수 있다. 덕분에 인덱싱도 빨라졌다. 초당 거의 3~4천 개의 이미지를 인덱싱할 수 있어서 수십, 수백만 장의 이미지를 보유한 고객들을 위해 프로젝트를 수행할 수 있는 역량을 확보했다”고 테일러는 말했습니다.

지프의 소프트웨어와 플래시블레이드(FlashBlade)는 사진 컬렉션을 인쇄 가능한 포토북으로 자동 전환해주는 인기 애플리케이션, Chatbooks를 위한 프로젝트에서 그 역량을 발휘했습니다. Chatbooks는 포토북에서 전혀 쓰지 않을지도 모르는 이미지들도 저장하기 때문에, 신속하고 정확하게 이를 여과해 삭제하는 “가비지(garbage)” 선별 방법이 필요했습니다.

Chatbooks의 한 데이터 과학자가 이미지들을 6주에 걸쳐 수동으로 라벨링 하여, 최종적으로 2만 개의 이미지로 구성된 데이터 세트를 만들었습니다. 목표는 서브 세트를 이용하여 수백만 개의 이미지로 구성된 Chatbooks의 데이터베이스에 있는 나머지 이미지들을 훈련시키는 것이었습니다.

플래시블레이드(FlashBlade)와 지프의 AI 어시스트 큐레이션을 사용해, 1,400만 개의 이미지 훈련 세트를 정제하고 24시간 이내에 새로운 후보 모델을 모두 훈련시킬 수 있었습니다. 테일러는 "플래시블레이드(FlashBlade)를 이용해 더 높은 처리 속도를 확보하고 데이터 셔플링[주1]을 제거하여 고객을 위한 가치 실현 시간이 단축되었다"고 말했습니다.

> " 저는 이를 '망각성'이라고 부릅니다. 플래시블레이드(FlashBlade)는 안정적이고 아주 간편하게 사용할 수 있기 때문에 있다는 그냥 두고 잊어버릴 수 있거든요."
>
> 벤 테일러(Ben Taylor), **최고 AI 책임자**

" [주1] 데이터 셔플링은 학습 모델의 견고함을 위한 AI 학습 과정의 한 부분입니다. 데이터 셔플링에는 2가지 방법이 있습니다. 첫 번째 데이터세트를 만들 때 셔플을 하는 것과, 두번째, GPU 학습을 위한 데이터세트를 만들지 않고 라벨된 데이터를 랜덤하게 학습하게 하는 것입니다. 학습 모델의 견고함을 위해서는 두 번째 방법을 권고하고 있으나 이 경우 I/O에 대한 중요한 역할을 하게 됩니다.

클라우드 같은 간편성과 안정성

플래시블레이드(FlashBlade)로 얻은 또 다른 혜택은 안정성과 관리 간소화입니다. 테일러는 "더 이상 데이터센터에 가서 고장난 드라이브를 교체하고 NFS 마운트의 다운 문제를 해결해야 할 필요가 없어졌다"며 "안정성은 워크로드 확장과 직결되기 때문에, 플래시블레이드(FlashBlade)의 안정성은 지프에 엄청난 혜택을 주었다"고 말했습니다.

"나는 플래시블레이드를 '망각성'이라고 부른다. 플래시블레이드(FlashBlade)는 안정적이고 아주 간편하게 사용할 수 있기 때문에 그냥 두고 잊어버릴 수 있기 때문이다."

곤잘레스는 플래시블레이드(FlashBlade)의 가장 큰 혜택은 데이터 과학자들의 생산성을 유지해준다는 것이라며, "과학자들은 하드웨어 디버깅이나 클라우드에서 데이터를 불러오고 내보내는데 시간을 낭비하는 게 아니라 실질적인 비즈니스 문제들을 해결하는데 집중할 수 있게 된다. 지프의 비즈니스에서 데이터 과학자들의 생산성이 가장 중요한 비용 요소이다. 그리고 이들이 흥미로운 작업에 계속 몰두할 수 있도록 만들어 주는 것이 바로 그들을 우리 팀의 일원으로 유지하는데 최고의 방법이다."라고 덧붙였습니다.

그는 플래시블레이드(FlashBlade)의 성능이 미치는 한 가지 간과하기 쉬운 중요한 영향을 언급했습니다. "AI 프로젝트의 경우, 많은 시도 후에야 비로소 성공하며, 항상 새로운 접근방식을 시도해보거나 정확도를 향상하기 위해 반복해야 한다. 그렇기 때문에, 2천만 개의 이미지로 모델을 훈련시킬 수 있다고 성공한 것은 아니다. 진정한 성공은 다른 데이터세트을 6~7회 학습을 수행할 수 있는 것이다. 플래시블레이드(FlashBlade)는 과거 2~3주가 소요되던 모델 훈련 시간을 적게는 24~48시간으로 줄여줌으로써 우리가 이를 수행 할 수 있도록 해준다"고 말했습니다.

더 많은 고객 프로젝트를 맡은 것 이외에도, 지프는 플래시블레이드(FlashBlade)를 내부 워크로드에도 활용하고 있습니다. 테일러는 "오픈소스 코드와 내부 데이터세트를 플래시블레이드(FlashBlade)로 로드할 수 있게 되었다. 이는 다른 스토리지 방식으로는 불가능했을 것이다. 덕분에 데이터 과학자들이 더 많이 반복하고 더 신속하게 학습할 수 있게 되었다"라고 말했습니다.

뿐만 아니라, 곤잘레스는 다수의 다양한 데이터세트와 워크로드를 플래시블레이드(FlashBlade)에 저장하여 GPU들을 지속적으로 구동할 수 있게 되었다고 말했습니다. 그는 "고객들을 위한 작업을 모두 처리하는 것 이외에도, 내부 연구 프로젝트와 다른 모델들에 데이터를 공급할 수 있게 되었다. 모든 것이 플래시블레이드(FlashBlade)에서 동시적으로 실행된다"고 설명했습니다.

지프는 실시간 분석이 AI의 다음 단계라고 생각합니다. 테일러는 "항상 데이터 확보, 훈련, 소비라는 AI 프로젝트의 기본 절차를 따라왔다"며 "플래시블레이드(FlashBlade) 플랫폼이 데이터를 이동시키는 속도와 AI 엔진을 통해 데이터를 스트리밍하는 동시에 그 결과를 모바일 사용자들까지 열람할 수 있게 할 가능성을 봤다. 이는 헬스케어, 콜 센터나 국가 보안 부문에서 큰 영향을 미칠 수 있을 것"이라고 말했습니다.

PURESTORAGE

korea@purestorage.com

www.purestorage.com/kr/customers

© 2018 Pure Storage, Inc. All rights reserved. Pure Storage, Pure1, Evergreen, FlashBlade, 및 P Logo는 Pure Storage, Inc.의 등록상표입니다.
모든 다른 등록 상표는 해당 소유업체의 재산입니다. ps_cs_ZIFF_01