



# NetApp 파일-객체 이중성과 AWS SageMaker를 활용한 클라우드 데이터 관리

NetApp artificial intelligence solutions

NetApp  
August 18, 2025

# 목차

NetApp 파일-객체 이중성과 AWS SageMaker를 활용한 클라우드 데이터 관리 .....	1
TR-4967: NetApp 파일-객체 이중성 및 AWS SageMaker를 활용한 클라우드 데이터 관리 .....	1
솔루션 기술 .....	1
사용 사례 요약 .....	2
데이터 과학자 및 기타 애플리케이션을 위한 데이터 이중성 .....	2
기술 요구 사항 .....	2
배포 절차 .....	3
Jupyter Notebooks에서 머신 러닝 검증 .....	17
결론 .....	29
추가 정보를 찾을 수 있는 곳 .....	29

# NetApp 파일-객체 이중성과 AWS SageMaker를 활용한 클라우드 데이터 관리

## TR-4967: NetApp 파일-객체 이중성 및 AWS SageMaker를 활용한 클라우드 데이터 관리

Karthikeyan Nagalingam, NetApp

데이터 과학자와 엔지니어는 종종 NFS 형식으로 저장된 데이터에 액세스해야 하지만, AWS SageMaker의 S3 프로토콜에서 직접 이 데이터에 액세스하는 것은 어려울 수 있습니다. AWS는 S3 버킷 액세스만 지원하기 때문입니다. 하지만 NetApp ONTAP NFS와 S3에 대한 이중 프로토콜 액세스를 활성화하여 솔루션을 제공합니다. 이 솔루션을 사용하면 데이터 과학자와 엔지니어가 NetApp Cloud Volumes ONTAP의 S3 버킷을 통해 AWS SageMaker 노트북에서 NFS 데이터에 액세스할 수 있습니다. 이 접근 방식을 사용하면 추가 소프트웨어가 필요 없이 NFS와 S3 모두에서 동일한 데이터에 쉽게 액세스하고 공유할 수 있습니다.

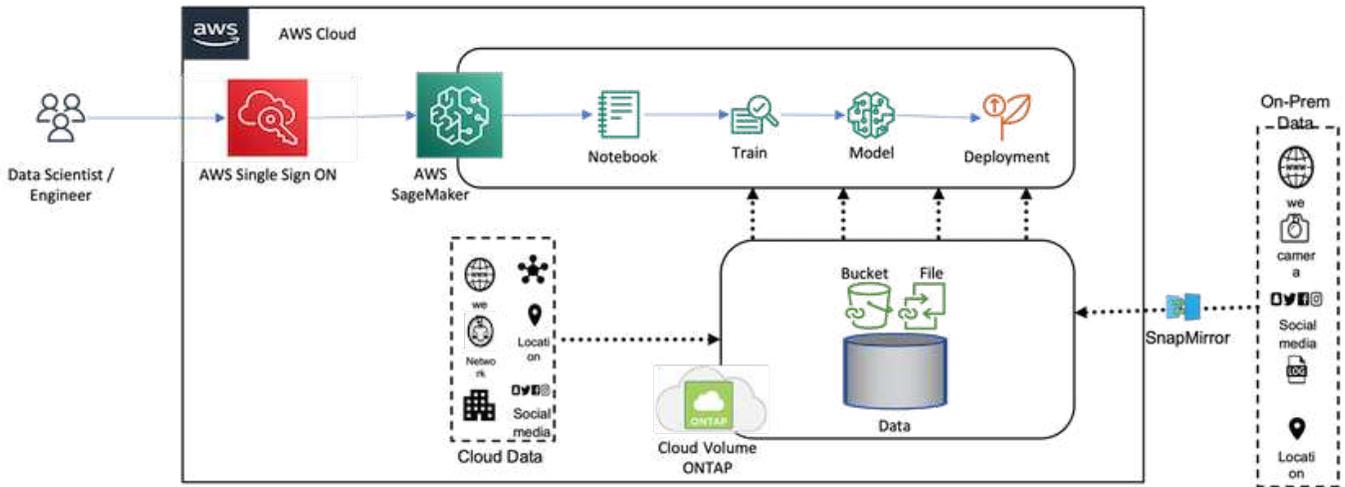
### 솔루션 기술

이 솔루션은 다음 기술을 활용합니다.

- **AWS SageMaker** 노트북. 개발자와 데이터 과학자에게 머신 러닝 기능을 제공하여 고품질 ML 모델을 효율적으로 만들고, 학습시키고, 배포할 수 있도록 지원합니다.
- \* NetApp BlueXP.\* 온프레미스뿐 아니라 AWS, Azure, Google Cloud에서도 스토리지를 검색, 배포, 운영할 수 있습니다. 데이터 손실, 사이버 위협, 계획되지 않은 중단으로부터 데이터를 보호하고 데이터 저장소와 인프라를 최적화합니다.
- \* NetApp Cloud Volumes ONTAP.\* AWS, Azure, Google Cloud에서 NFS, SMB/CIFS, iSCSI, S3 프로토콜을 갖춘 엔터프라이즈급 스토리지 볼륨을 제공하여 사용자가 클라우드에서 데이터에 액세스하고 관리할 수 있는 유연성을 높여줍니다.

BlueXP에서 생성된 NetApp Cloud Volumes ONTAP ML 데이터를 저장합니다.

다음 그림은 솔루션의 기술적 구성 요소를 보여줍니다.



## 사용 사례 요약

NFS와 S3의 이중 프로토콜 액세스에 대한 잠재적 사용 사례로는 머신 러닝과 데이터 과학 분야가 있습니다. 예를 들어, 데이터 과학자 팀이 AWS SageMaker를 사용하여 머신 러닝 프로젝트를 진행하는 경우, NFS 형식으로 저장된 데이터에 액세스해야 합니다. 하지만 다른 팀원과 협업하거나 S3를 사용하는 다른 애플리케이션과 통합하기 위해 S3 버킷을 통해 데이터에 액세스하고 공유해야 할 수도 있습니다.

NetApp Cloud Volumes ONTAP 활용함으로써 팀은 데이터를 단일 위치에 저장하고 NFS와 S3 프로토콜을 모두 사용하여 액세스할 수 있습니다. 데이터 과학자는 AWS SageMaker에서 직접 NFS 형식의 데이터에 액세스할 수 있으며, 다른 팀원이나 애플리케이션은 S3 버킷을 통해 동일한 데이터에 액세스할 수 있습니다.

이러한 접근 방식을 사용하면 추가 소프트웨어나 서로 다른 저장 솔루션 간의 데이터 마이그레이션이 필요 없이 쉽고 효율적으로 데이터에 액세스하고 공유할 수 있습니다. 또한 팀원 간의 워크플로와 협업이 간소화되어 머신 러닝 모델을 더 빠르고 효과적으로 개발할 수 있습니다.

## 데이터 과학자 및 기타 애플리케이션을 위한 데이터 이중성

데이터는 NFS에서 사용할 수 있으며 AWS SageMaker의 S3에서 액세스할 수 있습니다.

### 기술 요구 사항

데이터 이중성 사용 사례에는 NetApp BlueXP, NetApp Cloud Volumes ONTAP 및 AWS SageMaker Notebooks가 필요합니다.

### 소프트웨어 요구 사항

다음 표는 사용 사례를 구현하는 데 필요한 소프트웨어 구성 요소를 나열합니다.

소프트웨어	수량
BlueXP	1
NetApp Cloud Volumes ONTAP	1
AWS SageMaker 노트북	1

## 배포 절차

데이터 이중성 솔루션을 배포하려면 다음 작업이 필요합니다.

- BlueXP 커넥터
- NetApp Cloud Volumes ONTAP
- 머신러닝을 위한 데이터
- AWS 세이지메이커
- Jupyter Notebooks에서 검증된 머신 러닝

### BlueXP 커넥터

이 검증에서는 AWS를 사용했습니다. Azure와 Google Cloud에도 적용할 수 있습니다. AWS에서 BlueXP 커넥터를 생성하려면 다음 단계를 완료하세요.

1. 우리는 BlueXP의 mcarl-marketplace-subscription에 기반한 자격 증명을 사용했습니다.
2. 사용자 환경에 적합한 지역(예: us-east-1[N. Virginia])을 선택하고 인증 방법(예: 역할 가정 또는 AWS 키)을 선택합니다. 이 검증에서는 AWS 키를 사용합니다.
3. 커넥터의 이름을 제공하고 역할을 만듭니다.
4. 공용 IP가 필요한지 여부에 따라 VPC, 서브넷 또는 키 쌍과 같은 네트워크 세부 정보를 제공합니다.
5. 소스 유형에서 HTTP, HTTPS 또는 SSH 액세스와 같은 보안 그룹에 대한 세부 정보(어디서나 가능, IP 범위 정보 등)를 제공합니다.
6. BlueXP 커넥터를 검토하고 생성합니다.
7. AWS 콘솔에서 BlueXP EC2 인스턴스 상태가 실행 중인지 확인하고, 네트워킹 탭에서 IP 주소를 확인하세요.
8. BlueXP 포털에서 커넥터 사용자 인터페이스에 로그인하거나 브라우저에서 IP 주소를 사용하여 액세스할 수 있습니다.

### NetApp Cloud Volumes ONTAP

BlueXP에서 Cloud Volumes ONTAP 인스턴스를 생성하려면 다음 단계를 완료하세요.

1. 새로운 작업 환경을 만들고, 클라우드 공급자를 선택하고, Cloud Volumes ONTAP 인스턴스 유형(단일 CVO, HA 또는 ONTAP 용 Amazon FSx ONTAP 등)을 선택합니다.
2. Cloud Volumes ONTAP 클러스터 이름 및 자격 증명과 같은 세부 정보를 제공합니다. 이 검증에서 우리는 Cloud Volumes ONTAP 인스턴스를 생성했습니다. svm\_sagemaker\_cvo\_sn1 .
3. Cloud Volumes ONTAP에 필요한 서비스를 선택하세요. 이 검증에서는 모니터링만 하기로 했기 때문에 \*데이터 감지 및 규정 준수\*와 \*클라우드 서비스에 대한 백업\*을 비활성화했습니다.
4. 위치 및 연결 섹션에서 AWS 지역, VPC, 서브넷, 보안 그룹, SSH 인증 방법, 비밀번호 또는 키 쌍을 선택합니다.
5. 충전 방법을 선택하세요. 이 검증에는 전문가 모드를 사용했습니다.
6. **POC** 및 소규모 워크로드, 데이터베이스 및 애플리케이션 데이터 프로덕션 워크로드, 비용 효율적인 **DR** 또는 \*최고 성능 프로덕션 워크로드\*와 같은 미리 구성된 패키지를 선택할 수 있습니다. 이 검증에서는 \*Poc 및 소규모 워크로드\*를 선택합니다.
7. 특정 크기, 허용되는 프로토콜, 내보내기 옵션으로 볼륨을 생성합니다. 이 검증에서 우리는 볼륨이라는 것을

생성했습니다. vol1 .

8. 프로필 디스크 유형과 계층화 정책을 선택하세요. 이 검증 과정에서는 \*저장소 효율성\*과 \*범용 SSD - 동적 성능\*을 비활성화했습니다.
9. 마지막으로 Cloud Volumes ONTAP 인스턴스를 검토하고 생성합니다. 그런 다음 BlueXP 가 Cloud Volumes ONTAP 작업 환경을 생성하는 동안 15~20분 정도 기다립니다.
10. Duality 프로토콜을 활성화하려면 다음 매개변수를 구성하세요. Duality 프로토콜(NFS/S3)은 ONTAP 9부터 지원됩니다. 12.1 이상.
  - a. 이 검증에서 우리는 SVM을 생성했습니다. svm\_sagemaker\_cvo\_sn1 그리고 볼륨 vol1 .
  - b. SVM이 NFS 및 S3에 대한 프로토콜을 지원하는지 확인하세요. 그렇지 않은 경우 이를 지원하도록 SVM을 수정하세요.

```

sagemaker_cvo_sn1::> vserver show -vserver svm_sagemaker_cvo_sn1
                                Vserver: svm_sagemaker_cvo_sn1
                                Vserver Type: data
                                Vserver Subtype: default
                                Vserver UUID: 911065dd-a8bc-11ed-bc24-
e1c0f00ad86b
                                Root Volume:
svm_sagemaker_cvo_sn1_root
                                Aggregate: aggr1
                                NIS Domain: -
                                Root Volume Security Style: unix
                                LDAP Client: -
                                Default Volume Language Code: C.UTF-8
                                Snapshot Policy: default
                                Data Services: data-cifs, data-
flexcache,
                                                data-iscsi, data-nfs,
                                                data-nvme-tcp
                                Comment:
                                Quota Policy: default
                                List of Aggregates Assigned: aggr1
                                Limit on Maximum Number of Volumes allowed: unlimited
                                Vserver Admin State: running
                                Vserver Operational State: running
                                Vserver Operational State Stopped Reason: -
                                Allowed Protocols: nfs, cifs, fcp, iscsi,
ndmp, s3
                                Disallowed Protocols: nvme
                                Is Vserver with Infinite Volume: false
                                QoS Policy Group: -
                                Caching Policy Name: -
                                Config Lock: false
                                IPspace Name: Default
                                Foreground Process: -
                                Logical Space Reporting: true
                                Logical Space Enforcement: false
                                Default Anti_ransomware State of the Vserver's Volumes: disabled
                                Enable Analytics on New Volumes: false
                                Enable Activity Tracking on New Volumes: false

sagemaker_cvo_sn1::>

```

11. 필요한 경우 CA 인증서를 만들고 설치합니다.

12. 서비스 데이터 정책을 만듭니다.

```
sagemaker_cvo_sn1::*> network interface service-policy create -vserver
svm_sagemaker_cvo_sn1 -policy sagemaker_s3_nfs_policy -services data-
core,data-s3-server,data-nfs,data-flexcache
sagemaker_cvo_sn1::*> network interface create -vserver
svm_sagemaker_cvo_sn1 -lif svm_sagemaker_cvo_sn1_s3_lif -service-policy
sagemaker_s3_nfs_policy -home-node sagemaker_cvo_sn1-01 -address
172.30.10.41 -netmask 255.255.255.192
```

Warning: The configured failover-group has no valid failover targets for the LIF's failover-policy. To view the failover targets for a LIF, use the "network interface show -failover" command.

```
sagemaker_cvo_sn1::*>
```

```
sagemaker_cvo_sn1::*> network interface show
```

Logical Vserver Home	Status Interface	Network Admin/Oper	Current Address/Mask	Current Node	Is Port
sagemaker_cvo_sn1-01	cluster-mgmt	up/up	172.30.10.40/26	sagemaker_cvo_sn1-	e0a
true					
sagemaker_cvo_sn1-01	intercluster	up/up	172.30.10.48/26	sagemaker_cvo_sn1-	e0a
true					
sagemaker_cvo_sn1-01	sagemaker_cvo_sn1-01_mgmt1	up/up	172.30.10.58/26	sagemaker_cvo_sn1-	e0a
true					
svm_sagemaker_cvo_sn1-01	svm_sagemaker_cvo_sn1_data_lif	up/up	172.30.10.23/26	sagemaker_cvo_sn1-	e0a
true					
svm_sagemaker_cvo_sn1-01	svm_sagemaker_cvo_sn1_mgmt_lif	up/up	172.30.10.32/26	sagemaker_cvo_sn1-	e0a
true					
svm_sagemaker_cvo_sn1-01	svm_sagemaker_cvo_sn1_s3_lif	up/up	172.30.10.41/26	sagemaker_cvo_sn1-	

true

6 entries were displayed.

```
sagemaker_cvo_sn1::~*>
```

```
sagemaker_cvo_sn1::~*> vserver object-store-server create -vserver  
svm_sagemaker_cvo_sn1 -is-http-enabled true -object-store-server  
svm_sagemaker_cvo_s3_sn1 -is-https-enabled false  
sagemaker_cvo_sn1::~*> vserver object-store-server show
```

```
Vserver: svm_sagemaker_cvo_sn1
```

```
    Object Store Server Name: svm_sagemaker_cvo_s3_sn1
```

```
        Administrative State: up
```

```
            HTTP Enabled: true
```

```
    Listener Port For HTTP: 80
```

```
        HTTPS Enabled: false
```

```
    Secure Listener Port For HTTPS: 443
```

```
    Certificate for HTTPS Connections: -
```

```
        Default UNIX User: pcuser
```

```
    Default Windows User: -
```

```
        Comment:
```

```
sagemaker_cvo_sn1::~*>
```

13. 집계 세부 정보를 확인하세요.

```
sagemaker_cvo_sn1::*> aggr show
```

```
Aggregate      Size Available Used% State  #Vols  Nodes      RAID
Status
-----
-----
aggr0_sagemaker_cvo_sn1_01
      124.0GB   50.88GB   59% online    1 sagemaker_cvo_
raid0,
                                sn1-01
normal
aggr1      907.1GB   904.9GB   0% online    2 sagemaker_cvo_
raid0,
                                sn1-01
normal
2 entries were displayed.

sagemaker_cvo_sn1::*>
```

14. 사용자와 그룹을 만듭니다.

```

sagemaker_cvo_sn1::*> vserver object-store-server user create -vserver
svm_sagemaker_cvo_sn1 -user s3user

sagemaker_cvo_sn1::*> vserver object-store-server user show
Vserver      User          ID      Access Key      Secret Key
-----
svm_sagemaker_cvo_sn1
      root          0      -              -
      Comment: Root User
svm_sagemaker_cvo_sn1
      s3user        1      0ZNAX21JW5Q8AP80CQ2E
PpLs4gA9K0_2gPhuykkp014gBjcC9Rbi3QDX_6rr
2 entries were displayed.

sagemaker_cvo_sn1::*>

sagemaker_cvo_sn1::*> vserver object-store-server group create -name
s3group -users s3user -comment ""

sagemaker_cvo_sn1::*>
sagemaker_cvo_sn1::*> vserver object-store-server group delete -gid 1
-vserver svm_sagemaker_cvo_sn1

sagemaker_cvo_sn1::*> vserver object-store-server group create -name
s3group -users s3user -comment "" -policies FullAccess

sagemaker_cvo_sn1::*>

```

15. NFS 볼륨에 버킷을 생성합니다.

```
sagemaker_cvo_sn1::*> vservers object-store-server bucket create -bucket
ontapbucket1 -type nas -comment "" -vservers svm_sagemaker_cvo_sn1 -nas
-path /vol1
sagemaker_cvo_sn1::*> vservers object-store-server bucket show
Vserver      Bucket      Type      Volume      Size
Encryption  Role        NAS Path
-----
svm_sagemaker_cvo_sn1
      ontapbucket1  nas      vol1        -          false
-          /vol1
sagemaker_cvo_sn1::*>
```

## AWS 세이지메이커

AWS SageMaker에서 AWS Notebook을 만들려면 다음 단계를 완료하세요.

1. Notebook 인스턴스를 생성하는 사용자에게 AmazonSageMakerFullAccess IAM 정책이 있는지 또는 AmazonSageMakerFullAccess 권한이 있는 기존 그룹에 속해 있는지 확인하세요. 이 검증에서 사용자는 기존 그룹의 일부입니다.
2. 다음 정보를 제공하세요.
  - 노트북 인스턴스 이름.
  - 인스턴스 유형.
  - 플랫폼 식별자.
  - AmazonSageMakerFullAccess 권한이 있는 IAM 역할을 선택합니다.
  - 루트 접근 - 활성화.
  - 암호화 키 - 사용자 정의 암호화를 선택하지 않습니다.
  - 나머지 기본 옵션은 그대로 둡니다.
3. 이 검증에서 SageMaker 인스턴스 세부 정보는 다음과 같습니다.

Amazon SageMaker > Notebook instances > nkarthiksagemaker

### nkarthiksagemaker

Delete Stop Open Jupyter Open JupyterLab

Notebook instance settings Edit

Name	Status	Notebook instance type	Platform identifier
nkarthiksagemaker	<span style="color: green;">✔ InService</span>	ml.t2.medium	Amazon Linux 2, Jupyter Lab 3 (notebook-a2-v2)
ARN	Creation time	Elastic Inference	Minimum IMDS Version
arn:aws:sagemaker:us-east-1:210811600188:notebook-instance/nkarthiksagemaker	Feb 16, 2023 18:55 UTC	-	2
Lifecycle configuration	Last updated	Volume Size	
-	Mar 22, 2023 20:59 UTC	5GB EBS	

## Permissions and encryption

IAM role ARN

[arn:aws:iam::210811600188:role/SageMakerFullRole](#)

Root access

Enabled

Encryption key

## Network

Subnet(s)

[subnet-00f94558](#)

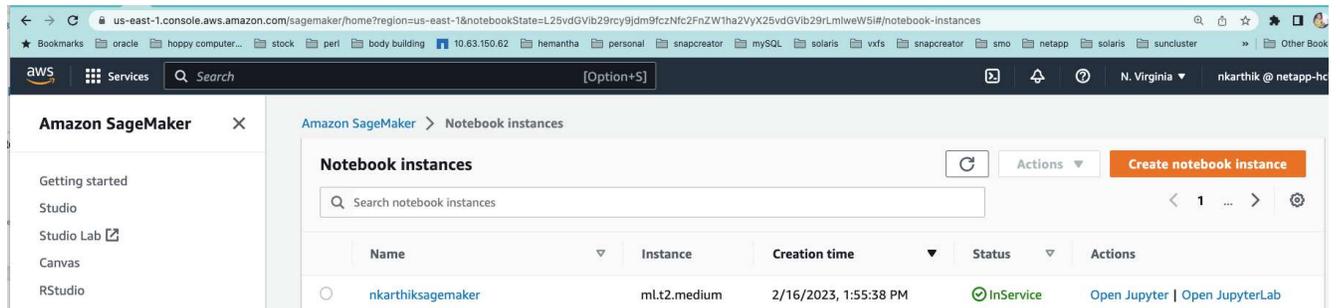
Security Group(s)

[sg-07111a8c16d67c81d](#)

Direct internet access

Enabled: [Learn more](#)

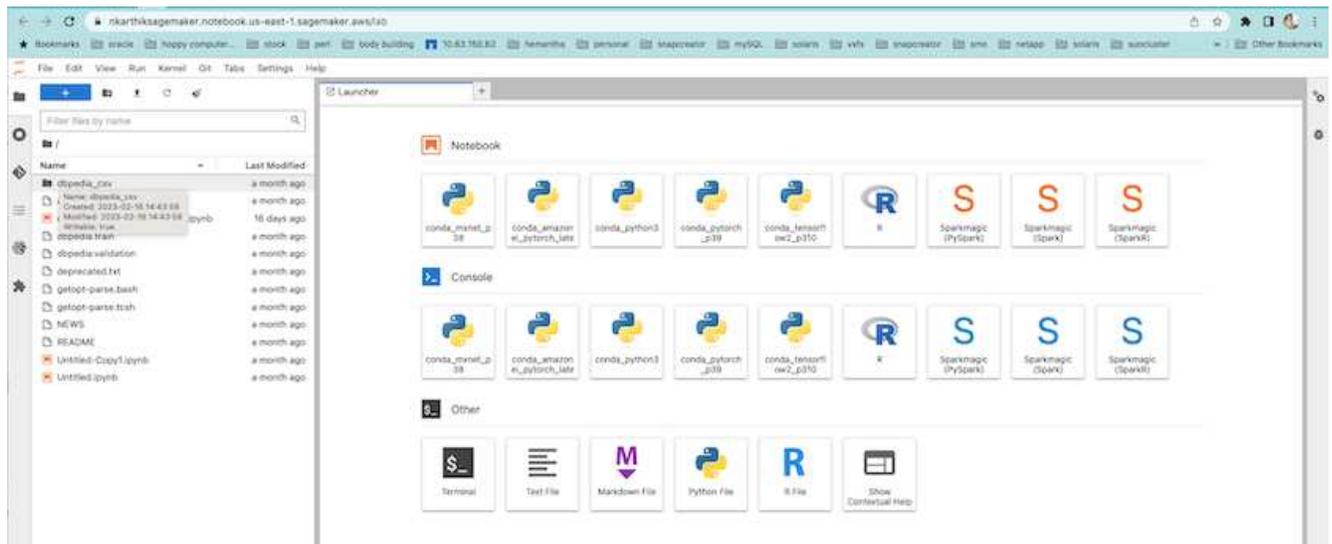
### 4. AWS Notebook을 시작합니다.



The screenshot shows the AWS SageMaker console interface. The main content area displays 'Notebook instances' with a search bar and a table of instances. The table has columns for Name, Instance, Creation time, Status, and Actions. One instance is listed: 'nkarthiksagemaker' with instance type 'ml.t2.medium', creation time '2/16/2023, 1:55:38 PM', and status 'InService'. The Actions column for this instance includes 'Open Jupyter' and 'Open JupyterLab'.

Name	Instance	Creation time	Status	Actions
nkarthiksagemaker	ml.t2.medium	2/16/2023, 1:55:38 PM	InService	Open Jupyter   Open JupyterLab

### 5. Jupyter 랩을 엽니다.



6. 터미널에 로그인하여 Cloud Volumes ONTAP 볼륨을 마운트합니다.

```
sh-4.2$ sudo mkdir /vol1; sudo mount -t nfs 172.30.10.41:/vol1 /vol1
sh-4.2$ df -h
```

Filesystem	Size	Used	Avail	Use%	Mounted on
devtmpfs	2.0G	0	2.0G	0%	/dev
tmpfs	2.0G	0	2.0G	0%	/dev/shm
tmpfs	2.0G	624K	2.0G	1%	/run
tmpfs	2.0G	0	2.0G	0%	/sys/fs/cgroup
/dev/xvda1	140G	114G	27G	82%	/
/dev/xvdf	4.8G	72K	4.6G	1%	/home/ec2-user/SageMaker
tmpfs	393M	0	393M	0%	/run/user/1001
tmpfs	393M	0	393M	0%	/run/user/1002
tmpfs	393M	0	393M	0%	/run/user/1000
172.30.10.41:/vol1	973M	189M	785M	20%	/vol1

```
sh-4.2$
```

7. AWS CLI 명령을 사용하여 Cloud Volumes ONTAP 볼륨에 생성된 버킷을 확인합니다.

```
sh-4.2$ aws configure --profile netapp
AWS Access Key ID [None]: 0ZNAX21JW5Q8AP80CQ2E
AWS Secret Access Key [None]: PpLs4gA9K0_2gPhuykkp014gBjcc9Rbi3QDX_6rr
Default region name [None]: us-east-1
Default output format [None]:
sh-4.2$

sh-4.2$ aws s3 ls --profile netapp --endpoint-url
2023-02-10 17:59:48 ontapbucket1

sh-4.2$ aws s3 ls --profile netapp --endpoint-url s3://ontapbucket1/

2023-02-10 18:46:44          4747 1
2023-02-10 18:48:32           96 setup.cfg

sh-4.2$
```

## 머신러닝을 위한 데이터

이 검증 과정에서 우리는 클라우드 소싱 커뮤니티 활동인 DBpedia의 데이터 세트를 사용하여 다양한 위키미디어 프로젝트에서 생성된 정보로부터 구조화된 콘텐츠를 추출했습니다.

1. DBpedia GitHub 위치에서 데이터를 다운로드하고 추출합니다. 이전 섹션에서 사용한 것과 동일한 터미널을 사용하세요.

```

sh-4.2$ wget
--2023-02-14 23:12:11--
Resolving github.com (github.com)... 140.82.113.3
Connecting to github.com (github.com)|140.82.113.3|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: [following]
--2023-02-14 23:12:11--
Resolving raw.githubusercontent.com (raw.githubusercontent.com)...
185.199.109.133, 185.199.110.133, 185.199.111.133, ...
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|185.199.109.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 68431223 (65M) [application/octet-stream]
Saving to: 'dbpedia_csv.tar.gz'

100%[=====
=====
=====>] 68,431,223  56.2MB/s  in 1.2s

2023-02-14 23:12:13 (56.2 MB/s) - 'dbpedia_csv.tar.gz' saved
[68431223/68431223]

sh-4.2$ tar -zxvf dbpedia_csv.tar.gz
dbpedia_csv/
dbpedia_csv/test.csv
dbpedia_csv/classes.txt
dbpedia_csv/train.csv
dbpedia_csv/readme.txt
sh-4.2$

```

2. 데이터를 Cloud Volumes ONTAP 위치로 복사한 다음 AWS CLI를 사용하여 S3 버킷에서 확인합니다.

```

sh-4.2$ df -h
Filesystem      Size  Used Avail Use% Mounted on
devtmpfs        2.0G   0  2.0G   0% /dev
tmpfs           2.0G   0  2.0G   0% /dev/shm
tmpfs           2.0G 628K  2.0G   1% /run
tmpfs           2.0G   0  2.0G   0% /sys/fs/cgroup
/dev/xvda1      140G 114G   27G  82% /
/dev/xvdf       4.8G  52K  4.6G   1% /home/ec2-user/SageMaker
tmpfs          393M   0  393M   0% /run/user/1002
tmpfs          393M   0  393M   0% /run/user/1001
tmpfs          393M   0  393M   0% /run/user/1000
172.30.10.41:/vol1 973M 384K  973M   1% /vol1
sh-4.2$ pwd
/home/ec2-user
sh-4.2$ cp -ra dbpedia_csv /vol1
sh-4.2$ aws s3 ls --profile netapp --endpoint-url s3://ontapbucket1/
                PRE dbpedia_csv/
2023-02-10 18:46:44          4747 1
2023-02-10 18:48:32           96 setup.cfg
sh-4.2$

```

### 3. 기본 검증을 수행하여 S3 버킷에서 읽기/쓰기 기능이 작동하는지 확인합니다.

```

sh-4.2$ aws s3 cp --profile netapp --endpoint-url /usr/share/doc/util-
linux-2.30.2 s3://ontapbucket1/ --recursive
upload: ../../usr/share/doc/util-linux-2.30.2/deprecated.txt to
s3://ontapbucket1/deprecated.txt
upload: ../../usr/share/doc/util-linux-2.30.2/getopt-parse.bash to
s3://ontapbucket1/getopt-parse.bash
upload: ../../usr/share/doc/util-linux-2.30.2/README to
s3://ontapbucket1/README
upload: ../../usr/share/doc/util-linux-2.30.2/getopt-parse.tcsh to
s3://ontapbucket1/getopt-parse.tcsh
upload: ../../usr/share/doc/util-linux-2.30.2/AUTHORS to
s3://ontapbucket1/AUTHORS
upload: ../../usr/share/doc/util-linux-2.30.2/NEWS to
s3://ontapbucket1/NEWS
sh-4.2$ aws s3 ls --profile netapp --endpoint-url
s3://ontapbucket1/s3://ontapbucket1/

An error occurred (InternalError) when calling the ListObjectsV2
operation: We encountered an internal error. Please try again.
sh-4.2$ aws s3 ls --profile netapp --endpoint-url s3://ontapbucket1/
                PRE dbpedia_csv/

```

```

2023-02-16 19:19:27      26774 AUTHORS
2023-02-16 19:19:27      72727 NEWS
2023-02-16 19:19:27       4493 README
2023-02-16 19:19:27       2825 deprecated.txt
2023-02-16 19:19:27       1590 getopt-parse.bash
2023-02-16 19:19:27       2245 getopt-parse.tcsh
sh-4.2$ ls -ltr /voll
total 132
drwxrwxr-x 2 ec2-user ec2-user  4096 Mar 29  2015 dbpedia_csv
-rw-r--r-- 1 nobody  nobody   2245 Apr 10 17:37 getopt-parse.tcsh
-rw-r--r-- 1 nobody  nobody   2825 Apr 10 17:37 deprecated.txt
-rw-r--r-- 1 nobody  nobody   4493 Apr 10 17:37 README
-rw-r--r-- 1 nobody  nobody   1590 Apr 10 17:37 getopt-parse.bash
-rw-r--r-- 1 nobody  nobody  26774 Apr 10 17:37 AUTHORS
-rw-r--r-- 1 nobody  nobody  72727 Apr 10 17:37 NEWS
sh-4.2$ ls -ltr /voll/dbpedia_csv/
total 192104
-rw----- 1 ec2-user ec2-user 174148970 Mar 28  2015 train.csv
-rw----- 1 ec2-user ec2-user  21775285 Mar 28  2015 test.csv
-rw----- 1 ec2-user ec2-user      146 Mar 28  2015 classes.txt
-rw-rw-r-- 1 ec2-user ec2-user   1758 Mar 29  2015 readme.txt
sh-4.2$ chmod -R 777 /voll/dbpedia_csv
sh-4.2$ ls -ltr /voll/dbpedia_csv/
total 192104
-rwxrwxrwx 1 ec2-user ec2-user 174148970 Mar 28  2015 train.csv
-rwxrwxrwx 1 ec2-user ec2-user  21775285 Mar 28  2015 test.csv
-rwxrwxrwx 1 ec2-user ec2-user      146 Mar 28  2015 classes.txt
-rwxrwxrwx 1 ec2-user ec2-user   1758 Mar 29  2015 readme.txt
sh-4.2$ aws s3 cp --profile netapp --endpoint-url http://172.30.2.248/
s3://ontapbucket1/ /tmp --recursive
download: s3://ontapbucket1/AUTHORS to ../../tmp/AUTHORS
download: s3://ontapbucket1/README to ../../tmp/README
download: s3://ontapbucket1/NEWS to ../../tmp/NEWS
download: s3://ontapbucket1/dbpedia_csv/classes.txt to
../../tmp/dbpedia_csv/classes.txt
download: s3://ontapbucket1/dbpedia_csv/readme.txt to
../../tmp/dbpedia_csv/readme.txt
download: s3://ontapbucket1/deprecated.txt to ../../tmp/deprecated.txt
download: s3://ontapbucket1/getopt-parse.bash to ../../tmp/getopt-
parse.bash
download: s3://ontapbucket1/getopt-parse.tcsh to ../../tmp/getopt-
parse.tcsh
download: s3://ontapbucket1/dbpedia_csv/test.csv to
../../tmp/dbpedia_csv/test.csv
download: s3://ontapbucket1/dbpedia_csv/train.csv to
../../tmp/dbpedia_csv/train.csv

```

```

sh-4.2$
sh-4.2$ aws s3 ls --profile netapp --endpoint-url s3://ontapbucket1/
                PRE dbpedia_csv/
2023-02-16 19:19:27      26774 AUTHORS
2023-02-16 19:19:27      72727 NEWS
2023-02-16 19:19:27      4493 README
2023-02-16 19:19:27      2825 deprecated.txt
2023-02-16 19:19:27      1590 getopt-parse.bash
2023-02-16 19:19:27      2245 getopt-parse.tcsh
sh-4.2$

```

## Jupyter Notebooks에서 머신 러닝 검증

다음 검증은 아래의 SageMaker BlazingText 예제를 사용하여 텍스트 분류를 통해 머신 러닝 모델을 빌드, 학습 및 배포하는 방법을 제공합니다.

1. boto3 및 SageMaker 패키지를 설치합니다.

```
In [1]: pip install --upgrade boto3 sagemaker
```

산출:

```

Looking in indexes: https://pypi.org/simple,
https://pip.repos.neuron.amazonaws.com
Requirement already satisfied: boto3 in /home/ec2-
user/anaconda3/envs/python3/lib/python3.10/site-packages (1.26.44)
Collecting boto3
  Downloading boto3-1.26.72-py3-none-any.whl (132 kB)
  _____
132.7/132.7 kB 14.6 MB/s eta 0: 00:00
Requirement already satisfied: sagemaker in /home/ec2-
user/anaconda3/envs/python3/lib/python3.10/site-packages (2.127.0)
Collecting sagemaker
  Downloading sagemaker-2.132.0.tar.gz (668 kB)
  _____
668.0/668.0 kB 12.3 MB/s eta 0:
00:0000:01
  Preparing metadata (setup.py) ... done
Collecting botocore<1.30.0,>=1.29.72
  Downloading botocore-1.29.72-py3-none-any.whl (10.4 MB)
  _____
10.4/10.4 MB 44.3 MB/s eta 0: 00:0000:010:01
Requirement already satisfied: s3transfer<0.7.0,>=0.6.0 in /home/ec2-
user/anaconda3/envs/python3/lib/python3.10/site-packages (from boto3)

```

(0.6.0)  
Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from boto3) (0.10.0)  
Requirement already satisfied: attrs<23,>=20.3.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (22.1.0)  
Requirement already satisfied: google-pasta in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (0.2.0)  
Requirement already satisfied: numpy<2.0,>=1.9.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (1.22.4)  
Requirement already satisfied: protobuf<4.0,>=3.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (3.20.3)  
Requirement already satisfied: protobuf3-to-dict<1.0,>=0.1.5 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (0.1.5)  
Requirement already satisfied: smdebug\_rulesconfig==1.0.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (1.0.1)  
Requirement already satisfied: importlib-metadata<5.0,>=1.4.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (4.13.0)  
Requirement already satisfied: packaging>=20.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (21.3)  
Requirement already satisfied: pandas in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (1.5.1)  
Requirement already satisfied: pathos in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (0.3.0)  
Requirement already satisfied: schema in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (0.7.5)  
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from botocore<1.30.0,>=1.29.72->boto3) (2.8.2)  
Requirement already satisfied: urllib3<1.27,>=1.25.4 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from botocore<1.30.0,>=1.2

```

9.72->boto3) (1.26.8) Requirement already satisfied: zipp>=0.5 in
/home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages
(from importlib-metadata<5.0,>=1.4.0->sagemaker) (3.10.0)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /home/ec2-
user/anaconda3/envs/python3/lib/python3.10/site-packages (from
packaging>=20.0->sagemaker) (3.0.9)
Requirement already satisfied: six in /home/ec2-
user/anaconda3/envs/python
3/lib/python3.10/site-packages (from protobuf3-to-dict<1.0,>=0.1.5-
>sagemaker) (1.16.0)
Requirement already satisfied: pytz>=2020.1 in /home/ec2-
user/anaconda3/envs/python3/lib/python3.10/site-packages (from pandas-
>sagemaker) (2022.5)
Requirement already satisfied: ppft>=1.7.6.6 in /home/ec2-
user/anaconda3/envs/python3/lib/python3.10/site-packages (from pathos-
>sagemaker) (1.7.6.6) Requirement already satisfied:
multiprocess>=0.70.14 in /home/ec2-user/anac
onda3/envs/python3/lib/python3.10/site-packages (from pathos->sagemaker)
(0.70.14)
Requirement already satisfied: dill>=0.3.6 in /home/ec2-
user/anaconda3/envs/python3/lib/python3.10/site-packages (from pathos-
>sagemaker) (0.3.6)
Requirement already satisfied: pox>=0.3.2 in /home/ec2-
user/anaconda3/envs/python3/lib/python3.10/site-packages (from pathos-
>sagemaker) (0.3.2) Requirement already satisfied: contextlib2>=0.5.5 in
/home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages
(from schema->sagemaker) (21.
6.0) Building wheels for collected packages: sagemaker
  Building wheel for sagemaker (setup.py) ... done
  Created wheel for sagemaker: filename=sagemaker-2.132.0-py2.py3-none-
any.whl size=905449
sha256=f6100a5dc95627f2e2a49824e38f0481459a27805ee19b5a06ec
83db0252fd41
  Stored in directory: /home/ec2-
user/.cache/pip/wheels/60/41/b6/482e7ab096
520df034fbf2dddd244a1d7ba0681b27ef45aa61
Successfully built sagemaker
Installing collected packages: botocore, boto3, sagemaker
  Attempting uninstall: botocore      Found existing installation:
botocore 1.24.19
    Uninstalling botocore-1.24.19:      Successfully uninstalled
botocore-1.24.19
  Attempting uninstall: boto3      Found existing installation: boto3
1.26.44
    Uninstalling boto3-1.26.44:
      Successfully uninstalled boto3-1.26.44

```

```
Attempting uninstall: sagemaker      Found existing installation:
sagemaker 2.127.0
  Uninstalling sagemaker-2.127.0:
    Successfully uninstalled sagemaker-2.127.0
ERROR: pip's dependency resolver does not currently take into account
all the packages that are installed. This behaviour is the source of
the following dependency conflicts.
awscli 1.27.44 requires botocore==1.29.44, but you have botocore 1.29.72
which is incompatible.
aiobotocore 2.0.1 requires botocore<1.22.9,>=1.22.8, but you have
botocore 1.29.72 which is incompatible. Successfully installed boto3-
1.26.72 botocore-1.29.72 sagemaker-2.132.0 Note: you may need to restart
the kernel to use updated packages.
```

2. 다음 단계에서는 데이터(dbpedia\_csv)은 s3 버킷에서 다운로드됩니다. ontapbucket1 머신 러닝에 사용되는 Jupyter Notebook 인스턴스로.

```

In [2]: import sagemaker
In [3]: from sagemaker import get_execution_role
In [4]:
import json
import boto3
sess = sagemaker.Session()
role = get_execution_role()
print(role)
bucket = "ontapbucket1"
print(bucket)
sess.s3_client = boto3.client('s3',region_name='',aws_access_key_id =
'0ZNAX21JW5Q8AP80CQ2E', aws_secret_access_key =
'PpLs4gA9K0_2gPhuykkp014gBjcC9Rbi3QDX_6rr',
                                use_ssl = False, endpoint_url =
'http://172.30.10.41',

config=boto3.session.Config(signature_version='s3v4',
s3={'addressing_style':'path'}) )
sess.s3_resource = boto3.resource('s3',region_name='',aws_access_key_id
= '0ZNAX21JW5Q8AP80CQ2E', aws_secret_access_key =
'PpLs4gA9K0_2gPhuykkp014gBjcC9Rbi3QDX_6rr',
                                use_ssl = False, endpoint_url =
'http://172.30.10.41',

config=boto3.session.Config(signature_version='s3v4',
s3={'addressing_style':'path'}) )
prefix = "blazingtext/supervised"
import os
my_bucket = sess.s3_resource.Bucket(bucket)
my_bucket = sess.s3_resource.Bucket(bucket)
#os.mkdir('dbpedia_csv')
for s3_object in my_bucket.objects.all():
    filename = s3_object.key
    # print(filename)
    # print(s3_object.key)
    my_bucket.download_file(s3_object.key, filename)

```

3. 다음 코드는 추론 중에 실제 클래스 이름을 검색하는 데 사용되는 정수 인덱스에서 클래스 레이블로의 매핑을 생성합니다.

```

index_to_label = {}
with open("dbpedia_csv/classes.txt") as f:
    for i,label in enumerate(f.readlines()):
        index_to_label[str(i + 1)] = label.strip()

```

출력에는 파일과 폴더가 나열됩니다. ontapbucket1 AWS SageMaker 머신 러닝 검증을 위한 데이터로 사용되는 버킷입니다.

```
arn:aws:iam::210811600188:role/SageMakerFullRole ontapbucket1
AUTHORS
AUTHORS
NEWS
NEWS
README README
dbpedia_csv/classes.txt dbpedia_csv/classes.txt dbpedia_csv/readme.txt
dbpedia_csv/readme.txt dbpedia_csv/test.csv dbpedia_csv/test.csv
dbpedia_csv/train.csv dbpedia_csv/train.csv deprecated.txt
deprecated.txt getopt-parse.bash getopt-parse.bash getopt-parse.tcsh
getopt-parse.tcsh
In [5]: ls
AUTHORS          deprecated.txt    getopt-parse.tcsh NEWS
Untitled.ipynb dbpedia_csv/    getopt-parse.bash lost+found/
README
In [6]: ls -l dbpedia_csv
total 191344
-rw-rw-r-- 1 ec2-user ec2-user      146 Feb 16 19:43 classes.txt
-rw-rw-r-- 1 ec2-user ec2-user     1758 Feb 16 19:43 readme.txt
-rw-rw-r-- 1 ec2-user ec2-user  21775285 Feb 16 19:43 test.csv
-rw-rw-r-- 1 ec2-user ec2-user 174148970 Feb 16 19:43 train.csv
```

4. 데이터 전처리 단계를 시작하여 훈련 데이터를 공백으로 구분된 토큰화된 텍스트 형식으로 전처리합니다. 이 텍스트는 BlazingText 알고리즘과 nltk 라이브러리에서 사용되어 DBPedia 데이터 세트의 입력 문장을 토큰화합니다. nltk 토큰라이저와 기타 라이브러리를 다운로드하세요. 그만큼 transform\_instance 각 데이터 인스턴스에 병렬로 적용하려면 Python multiprocessing 모듈을 사용합니다.

```
In [7]: from random import shuffle
import multiprocessing
from multiprocessing import Pool
import csv
import nltk
nltk.download("punkt")
def transform_instance(row):
    cur_row = []
    label = "__label__" + index_to_label [row[0]] # Prefix the index-ed
label with __label__
    cur_row.append (label)
    cur_row.extend(nltk.word_tokenize(row[1].lower ()))
    cur_row.extend(nltk.word_tokenize(row[2].lower ()))
    return cur_row
def preprocess(input_file, output_file, keep=1):
```

```

all_rows = []
with open(input_file,"r") as csvinfile:
    csv_reader = csv.reader(csvinfile, delimiter=",")
    for row in csv_reader:
        all_rows.append(row)
shuffle(all_rows)
all_rows = all_rows[: int(keep * len(all_rows))]
pool = Pool(processes=multiprocessing.cpu_count())
transformed_rows = pool.map(transform_instance, all_rows)
pool.close()
pool.join()
with open(output_file, "w") as csvoutfile:
    csv_writer = csv.writer (csvoutfile, delimiter=" ",
lineterminator="\n")
    csv_writer.writerows (transformed_rows)

# Preparing the training dataset
# since preprocessing the whole dataset might take a couple of minutes,
# we keep 20% of the training dataset for this demo.
# Set keep to 1 if you want to use the complete dataset
preprocess("dbpedia_csv/train.csv","dbpedia.train", keep=0.2)
# Preparing the validation dataset
preprocess("dbpedia_csv/test.csv","dbpedia.validation")
sess = sagemaker.Session()
role = get_execution_role()
print (role) # This is the role that sageMaker would use to leverage Aws
resources (S3, Cloudwatch) on your behalf
bucket = sess.default_bucket() # Replace with your own bucket name if
needed
print("default Bucket::: ")
print(bucket)

```

산출:

```

[nltk_data] Downloading package punkt to /home/ec2-user/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
arn:aws:iam::210811600188:role/SageMakerFullRole default Bucket:::
sagemaker-us-east-1-210811600188

```

5. 포맷된 학습 데이터 세트를 S3에 업로드하여 SageMaker에서 학습 작업을 실행할 수 있도록 합니다. 그런 다음 Python SDK를 사용하여 두 개의 파일을 버킷과 접두사 위치에 업로드합니다.

```

In [8]: %%time
train_channel = prefix + "/train"
validation_channel = prefix + "/validation"
sess.upload_data(path="dbpedia.train", bucket=bucket,
key_prefix=train_channel)
sess.upload_data(path="dbpedia.validation", bucket=bucket,
key_prefix=validation_channel)
s3_train_data = "s3://{}/{}".format(bucket, train_channel)
s3_validation_data = "s3://{}/{}".format(bucket, validation_channel)

```

산출:

```

CPU times: user 546 ms, sys: 163 ms, total: 709 ms
Wall time: 1.32 s

```

6. 모델 아티팩트가 로드되는 S3에 출력 위치를 설정하여 아티팩트가 알고리즘의 학습 작업의 출력이 될 수 있도록 합니다. 생성하다 `sageMaker.estimator.Estimator` 훈련 작업을 시작하기 위한 객체입니다.

```

In [9]: s3_output_location = "s3://{}/{}/output".format(bucket, prefix)
In [10]: region_name = boto3.Session().region_name
In [11]: container =
sagemaker.amazon.amazon_estimator.get_image_uri(region_name,
"blazingtext", "latest")
print("Using SageMaker BlazingText container: {} ({})"
.format(container,
region_name))

```

산출:

```

The method get_image_uri has been renamed in sagemaker>=2.
See: https://sagemaker.readthedocs.io/en/stable/v2.html for details.
Defaulting to the only supported framework/algorithm version: 1.
Ignoring f ramework/algorithm version: latest.
Using SageMaker BlazingText container: 811284229777.dkr.ecr.us-east-1.
amazonaws.com/blazingtext:1 (us-east-1)

```

7. SageMaker 정의 `Estimator` DBPedia 데이터세트에서 `c4.4xlarge` 인스턴스의 감독 모드를 사용하여 텍스트 분류를 훈련하기 위한 리소스 구성과 하이퍼파라미터를 사용합니다.

```

In [12]: bt_model = sagemaker.estimator.Estimator(
    container,
    role,
    instance_count=1,
    instance_type="ml.c4.4xlarge",
    volume_size=30,
    max_run=360000,
    input_mode="File",
    output_path=s3_output_location,
    hyperparameters={
        "mode": "supervised",
        "epochs": 1,
        "min_count": 2,
        "learning_rate": 0.05,
        "vector_dim": 10,
        "early_stopping": True,
        "patience": 4,
        "min_epochs": 5,
        "word_ngrams": 2,
    },
)

```

8. 데이터 채널과 알고리즘 간의 핸드셰이크를 준비합니다. 이를 위해 다음을 생성합니다.  
 sagemaker.session.s3\_input 데이터 채널의 객체를 알고리즘이 사용할 수 있도록 사전에 보관합니다.

```

In [13]: train_data = sagemaker.inputs.TrainingInput(
    s3_train_data,
    distribution="FullyReplicated",
    content_type="text/plain",
    s3_data_type="S3Prefix",
)
validation_data = sagemaker.inputs.TrainingInput(
    s3_validation_data,
    distribution="FullyReplicated",
    content_type="text/plain",
    s3_data_type="S3Prefix",
)
data_channels = {"train": train_data, "validation": validation_data}

```

9. 작업이 완료되면 작업 완료 메시지가 나타납니다. 훈련된 모델은 설정된 S3 버킷에서 찾을 수 있습니다.  
 output\_path 추정자에서.

```

In [14]: bt_model.fit(inputs=data_channels, logs=True)

```

산출:

```
INFO:sagemaker:Creating training-job with name: blazingtext-2023-02-16-20-3
7-30-748
2023-02-16 20:37:30 Starting - Starting the training job.....
2023-02-16 20:38:09 Starting - Preparing the instances for
training.....
2023-02-16 20:39:24 Downloading - Downloading input data
2023-02-16 20:39:24 Training - Training image download completed.
Training in progress... Arguments: train
[02/16/2023 20:39:41 WARNING 140279908747072] Loggers have already been
set up. [02/16/2023 20:39:41 WARNING 140279908747072] Loggers have
already been set up.
[02/16/2023 20:39:41 INFO 140279908747072] nvidia-smi took:
0.0251793861389
16016 secs to identify 0 gpus
[02/16/2023 20:39:41 INFO 140279908747072] Running single machine CPU
Blazi ngText training using supervised mode.
Number of CPU sockets found in instance is 1
[02/16/2023 20:39:41 INFO 140279908747072] Processing
/opt/ml/input/data/tr ain/dbpedia.train . File size: 35.0693244934082 MB
[02/16/2023 20:39:41 INFO 140279908747072] Processing
/opt/ml/input/data/va lidation/dbpedia.validation . File size:
21.887572288513184 MB
Read 6M words
Number of words: 149301
Loading validation data from
/opt/ml/input/data/validation/dbpedia.validati on
Loaded validation data.
----- End of epoch: 1 ##### Alpha: 0.0000 Progress: 100.00%
Million Words/sec: 10.39 ##### Training finished.
Average throughput in Million words/sec: 10.39
Total training time in seconds: 0.60
#train_accuracy: 0.7223
Number of train examples: 112000
#validation_accuracy: 0.7205
Number of validation examples: 70000
2023-02-16 20:39:55 Uploading - Uploading generated training model
2023-02-16 20:40:11 Completed - Training job completed
Training seconds: 68
Billable seconds: 68
```

10. 학습이 완료되면 학습된 모델을 Amazon SageMaker 실시간 호스팅 엔드포인트로 배포하여 예측을 수행합니다.

```
In [15]: from sagemaker.serializers import JSONSerializer
        text_classifier = bt_model.deploy(
            initial_instance_count=1, instance_type="ml.m4.xlarge",
            serializer=JSONS
        )
```

산출:

```
INFO:sagemaker:Creating model with name: blazingtext-2023-02-16-20-41-33-10
0
INFO:sagemaker:Creating endpoint-config with name blazingtext-2023-02-16-20-41-33-100
INFO:sagemaker:Creating endpoint with name blazingtext-2023-02-16-20-41-33-100
-----!
```

```
In [16]: sentences = [
        "Convair was an american aircraft manufacturing company which later expanded into rockets and spacecraft.",
        "Berwick secondary college is situated in the outer melbourne metropolitan suburb of berwick .",
    ]
# using the same nltk tokenizer that we used during data preparation for training
tokenized_sentences = [" ".join(nltk.word_tokenize(sent)) for sent in sentences]
payload = {"instances": tokenized_sentences} response = text_classifier.predict(payload)
predictions = json.loads(response)
print(json.dumps(predictions, indent=2))
```

```
[
  {
    "label": [
      "__label__Artist"
    ],
    "prob": [
      0.4090951681137085
    ]
  },
  {
    "label": [
      "__label__EducationalInstitution"
    ],
    "prob": [
      0.49466073513031006
    ]
  }
]
```

11. 기본적으로 모델은 확률이 가장 높은 예측 하나를 반환합니다. 상단을 검색하려면 `k` 예측, 설정 `k` 구성 파일에서.

```
In [17]: payload = {"instances": tokenized_sentences, "configuration":
{"k": 2}}
response = text_classifier.predict(payload)

predictions = json.loads(response)
print(json.dumps(predictions, indent=2))
```

```
[
  {
    "label": [
      "__label__Artist",
      "__label__MeanOfTransportation"
    ],
    "prob": [
      0.4090951681137085,
      0.26930734515190125
    ]
  },
  {
    "label": [
      "__label__EducationalInstitution",
      "__label__Building"
    ],
    "prob": [
      0.49466073513031006,
      0.15817692875862122
    ]
  }
]
```

12. 노트북을 닫기 전에 엔드포인트를 삭제하세요.

```
In [18]: sess.delete_endpoint(text_classifier.endpoint)
WARNING:sagemaker.deprecations:The endpoint attribute has been renamed
in sagemaker>=2.
See: https://sagemaker.readthedocs.io/en/stable/v2.html for details.
INFO:sagemaker:Deleting endpoint with name: blazingtext-2023-02-16-20-
41-33
-100
```

## 결론

이러한 검증을 기반으로 데이터 과학자와 엔지니어는 NetApp Cloud Volumes ONTAP 의 S3 버킷을 통해 AWS SageMaker Jupyter Notebooks에서 NFS 데이터에 액세스할 수 있습니다. 이 접근 방식을 사용하면 추가 소프트웨어가 필요 없이 NFS와 S3 모두에서 동일한 데이터에 쉽게 액세스하고 공유할 수 있습니다.

추가 정보를 찾을 수 있는 곳

이 문서에 설명된 정보에 대해 자세히 알아보려면 다음 문서 및/또는 웹사이트를 검토하세요.

- SageMaker BlazingText를 사용한 텍스트 분류
- S3 객체 스토리지에 대한 ONTAP 버전 지원

["https://docs.netapp.com/us-en/ontap/s3-config/ontap-version-support-s3-concept.html"](https://docs.netapp.com/us-en/ontap/s3-config/ontap-version-support-s3-concept.html)

## 저작권 정보

Copyright © 2025 NetApp, Inc. All Rights Reserved. 미국에서 인쇄된 본 문서의 어떠한 부분도 저작권 소유자의 사전 서면 승인 없이는 어떠한 형식이나 수단(복사, 녹음, 녹화 또는 전자 검색 시스템에 저장하는 것을 비롯한 그래픽, 전자적 또는 기계적 방법)으로도 복제될 수 없습니다.

NetApp이 저작권을 가진 자료에 있는 소프트웨어에는 아래의 라이선스와 고지사항이 적용됩니다.

본 소프트웨어는 NetApp에 의해 '있는 그대로' 제공되며 상품성 및 특정 목적에의 적합성에 대한 명시적 또는 묵시적 보증을 포함하여(이에 제한되지 않음) 어떠한 보증도 하지 않습니다. NetApp은 대체품 또는 대체 서비스의 조달, 사용 불능, 데이터 손실, 이익 손실, 영업 중단을 포함하여(이에 국한되지 않음), 이 소프트웨어의 사용으로 인해 발생하는 모든 직접 및 간접 손해, 우발적 손해, 특별 손해, 징벌적 손해, 결과적 손해의 발생에 대하여 그 발생 이유, 책임론, 계약 여부, 엄격한 책임, 불법 행위(과실 또는 그렇지 않은 경우)와 관계없이 어떠한 책임도 지지 않으며, 이와 같은 손실의 발생 가능성이 통지되었다 하더라도 마찬가지입니다.

NetApp은 본 문서에 설명된 제품을 언제든지 예고 없이 변경할 권리를 보유합니다. NetApp은 NetApp의 명시적인 서면 동의를 받은 경우를 제외하고 본 문서에 설명된 제품을 사용하여 발생하는 어떠한 문제에도 책임을 지지 않습니다. 본 제품의 사용 또는 구매의 경우 NetApp에서는 어떠한 특허권, 상표권 또는 기타 지적 재산권이 적용되는 라이선스도 제공하지 않습니다.

본 설명서에 설명된 제품은 하나 이상의 미국 특허, 해외 특허 또는 출원 중인 특허로 보호됩니다.

제한적 권리 표시: 정부에 의한 사용, 복제 또는 공개에는 DFARS 252.227-7013(2014년 2월) 및 FAR 52.227-19(2007년 12월)의 기술 데이터-비상업적 품목에 대한 권리(Rights in Technical Data -Noncommercial Items) 조항의 하위 조항 (b)(3)에 설명된 제한사항이 적용됩니다.

여기에 포함된 데이터는 상업용 제품 및/또는 상업용 서비스(FAR 2.101에 정의)에 해당하며 NetApp, Inc.의 독점 자산입니다. 본 계약에 따라 제공되는 모든 NetApp 기술 데이터 및 컴퓨터 소프트웨어는 본질적으로 상업용이며 개인 비용만으로 개발되었습니다. 미국 정부는 데이터가 제공된 미국 계약과 관련하여 해당 계약을 지원하는 데에만 데이터에 대한 전 세계적으로 비독점적이고 양도할 수 없으며 재사용이 불가능하며 취소 불가능한 라이선스를 제한적으로 가집니다. 여기에 제공된 경우를 제외하고 NetApp, Inc.의 사전 서면 승인 없이는 이 데이터를 사용, 공개, 재생산, 수정, 수행 또는 표시할 수 없습니다. 미국 국방부에 대한 정부 라이선스는 DFARS 조항 252.227-7015(b)(2014년 2월)에 명시된 권한으로 제한됩니다.

## 상표 정보

NETAPP, NETAPP 로고 및 <http://www.netapp.com/TM>에 나열된 마크는 NetApp, Inc.의 상표입니다. 기타 회사 및 제품 이름은 해당 소유자의 상표일 수 있습니다.