



Cloud-Datenmanagement mit NetApp File-Object Duality und AWS SageMaker

NetApp artificial intelligence solutions

NetApp
December 04, 2025

Inhalt

- Cloud-Datenmanagement mit NetApp File-Object Duality und AWS SageMaker 1
 - TR-4967: Cloud-Datenverwaltung mit NetApp File-Object Duality und AWS SageMaker 1
- Lösungstechnologie 1
 - Zusammenfassung des Anwendungsfalls 2
- Datendualität für Datenwissenschaftler und andere Anwendungen 2
 - Technologieanforderungen 2
 - Bereitstellungsverfahren 3
 - Validieren Sie maschinelles Lernen aus Jupyter Notebooks 17
- Abschluss 29
 - Wo Sie weitere Informationen finden 30

Cloud-Datenmanagement mit NetApp File-Object Duality und AWS SageMaker

TR-4967: Cloud-Datenverwaltung mit NetApp File-Object Duality und AWS SageMaker

Karthikeyan Nagalingam, NetApp

Datenwissenschaftler und -ingenieure müssen häufig auf im NFS-Format gespeicherte Daten zugreifen. Der direkte Zugriff auf diese Daten über das S3-Protokoll in AWS SageMaker kann jedoch eine Herausforderung darstellen, da AWS nur den Zugriff auf S3-Buckets unterstützt. NetApp ONTAP bietet jedoch eine Lösung, indem es den Dual-Protokoll-Zugriff für NFS und S3 ermöglicht. Mit dieser Lösung können Datenwissenschaftler und Ingenieure über S3-Buckets von NetApp Cloud Volumes ONTAP auf NFS-Daten von AWS SageMaker-Notebooks zugreifen. Dieser Ansatz ermöglicht den einfachen Zugriff und die gemeinsame Nutzung derselben Daten sowohl von NFS als auch von S3, ohne dass zusätzliche Software erforderlich ist.

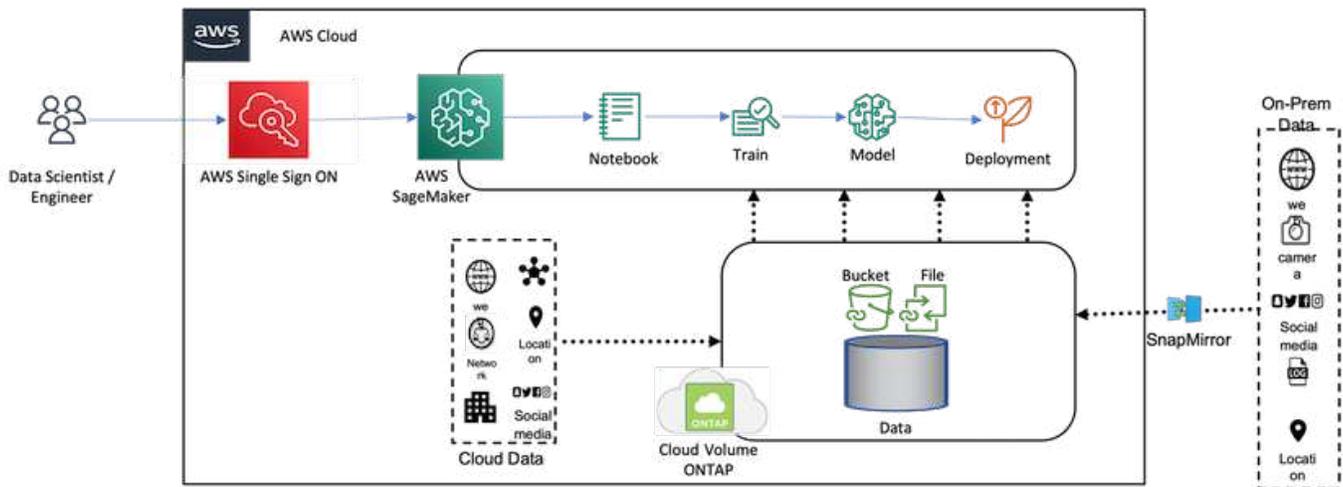
Lösungstechnologie

Diese Lösung nutzt die folgenden Technologien:

- **AWS SageMaker-Notizbuch.** Bietet Entwicklern und Datenwissenschaftlern Machine-Learning-Funktionen zum effizienten Erstellen, Trainieren und Bereitstellen hochwertiger ML-Modelle.
- * NetApp BlueXP.* Ermöglicht die Erkennung, Bereitstellung und den Betrieb von Speicher vor Ort sowie auf AWS, Azure und Google Cloud. Es bietet Datenschutz vor Datenverlust, Cyberbedrohungen und ungeplanten Ausfällen und optimiert die Datenspeicherung und Infrastruktur.
- * NetApp Cloud Volumes ONTAP.* Bietet Speichervolumen der Unternehmensklasse mit NFS-, SMB/CIFS-, iSCSI- und S3-Protokollen auf AWS, Azure und Google Cloud und bietet Benutzern so mehr Flexibilität beim Zugriff auf und der Verwaltung ihrer Daten in der Cloud.

Aus BlueXP erstelltes NetApp Cloud Volumes ONTAP zum Speichern von ML-Daten.

Die folgende Abbildung zeigt die technischen Komponenten der Lösung.



Zusammenfassung des Anwendungsfalls

Ein potenzieller Anwendungsfall für den Dualprotokollzugriff von NFS und S3 liegt in den Bereichen maschinelles Lernen und Datenwissenschaft. Beispielsweise könnte ein Team von Datenwissenschaftlern an einem Machine-Learning-Projekt mit AWS SageMaker arbeiten, das Zugriff auf im NFS-Format gespeicherte Daten erfordert. Möglicherweise müssen die Daten jedoch auch über S3-Buckets abgerufen und freigegeben werden, um mit anderen Teammitgliedern zusammenzuarbeiten oder sie in andere Anwendungen zu integrieren, die S3 verwenden.

Durch die Nutzung von NetApp Cloud Volumes ONTAP kann das Team seine Daten an einem einzigen Ort speichern und darauf sowohl mit NFS- als auch mit S3-Protokollen zugreifen. Die Datenwissenschaftler können direkt von AWS SageMaker auf die Daten im NFS-Format zugreifen, während andere Teammitglieder oder Anwendungen über S3-Buckets auf dieselben Daten zugreifen können.

Dieser Ansatz ermöglicht einen einfachen und effizienten Zugriff auf die Daten und deren gemeinsame Nutzung, ohne dass zusätzliche Software oder eine Datenmigration zwischen verschiedenen Speicherlösungen erforderlich ist. Darüber hinaus ermöglicht es einen optimierten Arbeitsablauf und eine bessere Zusammenarbeit zwischen den Teammitgliedern, was zu einer schnelleren und effektiveren Entwicklung von Modellen für maschinelles Lernen führt.

Datendualität für Datenwissenschaftler und andere Anwendungen

Die Daten sind in NFS verfügbar und können über S3 von AWS SageMaker abgerufen werden.

Technologieanforderungen

Sie benötigen NetApp BlueXP, NetApp Cloud Volumes ONTAP und AWS SageMaker Notebooks für den Anwendungsfall der Datendualität.

Softwareanforderungen

In der folgenden Tabelle sind die Softwarekomponenten aufgelistet, die zur Umsetzung des Anwendungsfalls erforderlich sind.

Software	Menge
BlueXP	1
NetApp Cloud Volumes ONTAP	1
AWS SageMaker-Notebook	1

Bereitstellungsverfahren

Die Bereitstellung der Datendualitätslösung umfasst die folgenden Aufgaben:

- BlueXP -Anschluss
- NetApp Cloud Volumes ONTAP
- Daten für maschinelles Lernen
- AWS SageMaker
- Validiertes maschinelles Lernen von Jupyter Notebooks

BlueXP -Anschluss

Bei dieser Validierung haben wir AWS verwendet. Es ist auch für Azure und Google Cloud anwendbar. Führen Sie die folgenden Schritte aus, um einen BlueXP Connector in AWS zu erstellen:

1. Wir haben die Anmeldeinformationen basierend auf dem mcarl-marketplace-subscription in BlueXP verwendet.
2. Wählen Sie die für Ihre Umgebung geeignete Region (z. B. us-east-1 [N. Virginia]) und wählen Sie die Authentifizierungsmethode (z. B. Rolle übernehmen oder AWS-Schlüssel). Bei dieser Validierung verwenden wir AWS-Schlüssel.
3. Geben Sie den Namen des Connectors an und erstellen Sie eine Rolle.
4. Geben Sie die Netzwerkdetails wie VPC, Subnetz oder Schlüsselpaar an, je nachdem, ob Sie eine öffentliche IP benötigen oder nicht.
5. Geben Sie die Details für die Sicherheitsgruppe an, z. B. HTTP-, HTTPS- oder SSH-Zugriff vom Quelltyp, z. B. Informationen zu einem beliebigen Ort und dem IP-Bereich.
6. Überprüfen und erstellen Sie den BlueXP Connector.
7. Stellen Sie sicher, dass der Status der BlueXP EC2-Instanz in der AWS-Konsole ausgeführt wird, und überprüfen Sie die IP-Adresse auf der Registerkarte **Netzwerk**.
8. Melden Sie sich über das BlueXP -Portal bei der Connector-Benutzeroberfläche an, oder verwenden Sie die IP-Adresse für den Zugriff über den Browser.

NetApp Cloud Volumes ONTAP

Führen Sie die folgenden Schritte aus, um eine Cloud Volumes ONTAP Instanz in BlueXP zu erstellen:

1. Erstellen Sie eine neue Arbeitsumgebung, wählen Sie den Cloud-Anbieter und den Typ der Cloud Volumes ONTAP Instanz aus (z. B. Single-CVO, HA oder Amazon FSx ONTAP für ONTAP).
2. Geben Sie Details wie den Namen und die Anmeldeinformationen des Cloud Volumes ONTAP Clusters an. Bei dieser Validierung haben wir eine Cloud Volumes ONTAP Instanz namens `svm_sagemaker_cvo_sn1`.
3. Wählen Sie die für Cloud Volumes ONTAP benötigten Dienste aus. Bei dieser Validierung haben wir uns

für die reine Überwachung entschieden und daher **Data Sense & Compliance** und **Backup to Cloud Services** deaktiviert.

4. Wählen Sie im Abschnitt **Standort und Konnektivität** die AWS-Region, VPC, das Subnetz, die Sicherheitsgruppe, die SSH-Authentifizierungsmethode und entweder ein Kennwort oder ein Schlüsselpaar aus.
5. Wählen Sie die Lademethode. Für diese Validierung haben wir **Professional** verwendet.
6. Sie können ein vorkonfiguriertes Paket auswählen, beispielsweise **POC und kleine Workloads, Produktions-Workloads für Datenbanken und Anwendungsdaten, Kostengünstige Notfallwiederherstellung** oder **Produktions-Workloads mit höchster Leistung**. Bei dieser Validierung wählen wir **Poc und kleine Workloads**.
7. Erstellen Sie ein Volume mit einer bestimmten Größe, zulässigen Protokollen und Exportoptionen. Bei dieser Validierung haben wir ein Volume namens `vol1` .
8. Wählen Sie einen Profildatenträgertyp und eine Tiering-Richtlinie. Bei dieser Validierung haben wir **Speichereffizienz** und **Allzweck-SSD – Dynamische Leistung** deaktiviert.
9. Überprüfen und erstellen Sie abschließend die Cloud Volumes ONTAP -Instanz. Warten Sie dann 15–20 Minuten, bis BlueXP die Cloud Volumes ONTAP Arbeitsumgebung erstellt hat.
10. Konfigurieren Sie die folgenden Parameter, um das Duality-Protokoll zu aktivieren. Das Duality-Protokoll (NFS/S3) wird ab ONTAP 9 unterstützt. 12.1 und höher.
 - a. Bei dieser Validierung haben wir eine SVM namens `svm_sagemaker_cvo_sn1` und Lautstärke `vol1` .
 - b. Überprüfen Sie, ob die SVM die Protokollunterstützung für NFS und S3 hat. Wenn nicht, ändern Sie die SVM, um sie zu unterstützen.

```

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. Erstellen und installieren Sie bei Bedarf ein CA-Zertifikat.

12. Erstellen Sie eine Servicedatenrichtlinie.

```
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-	

01

e0a

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. Überprüfen Sie die Gesamtdetails.

```
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,
normal
                                sn1-01
aggr1      907.1GB   904.9GB   0% online    2 sagemaker_cvo_
raid0,
normal
                                sn1-01
2 entries were displayed.

sagemaker_cvo_sn1::*>
```

14. Erstellen Sie einen Benutzer und eine Gruppe.

```

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

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

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

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

sagemaker_cvo_sn1::~*>

```

15. Erstellen Sie einen Bucket auf dem NFS-Volume.

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

AWS SageMaker

Führen Sie die folgenden Schritte aus, um ein AWS-Notebook aus AWS SageMaker zu erstellen:

1. Stellen Sie sicher, dass der Benutzer, der die Notebook-Instanz erstellt, über eine AmazonSageMakerFullAccess-IAM-Richtlinie verfügt oder Teil einer vorhandenen Gruppe mit AmazonSageMakerFullAccess-Rechten ist. Bei dieser Validierung ist der Benutzer Teil einer bestehenden Gruppe.
2. Geben Sie die folgenden Informationen an:
 - Name der Notebook-Instanz.
 - Instanztyp.
 - Plattformkennung.
 - Wählen Sie die IAM-Rolle aus, die über AmazonSageMakerFullAccess-Rechte verfügt.
 - Root-Zugriff – aktivieren.
 - Verschlüsselungsschlüssel – Wählen Sie keine benutzerdefinierte Verschlüsselung.
 - Behalten Sie die verbleibenden Standardoptionen bei.
3. Bei dieser Validierung lauten die Details der SageMaker-Instanz wie folgt:

Amazon SageMaker > Notebook instances > nkarthiksagemaker

nkarthiksagemaker Delete Stop Open Jupyter Open JupyterLab

Notebook instance settings Edit

Name	Status	Notebook instance type	Platform identifier
nkarthiksagemaker	InService	ml.t2.medium	Amazon Linux 2, Jupyter Lab 3 (notebook-ml2-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](#)

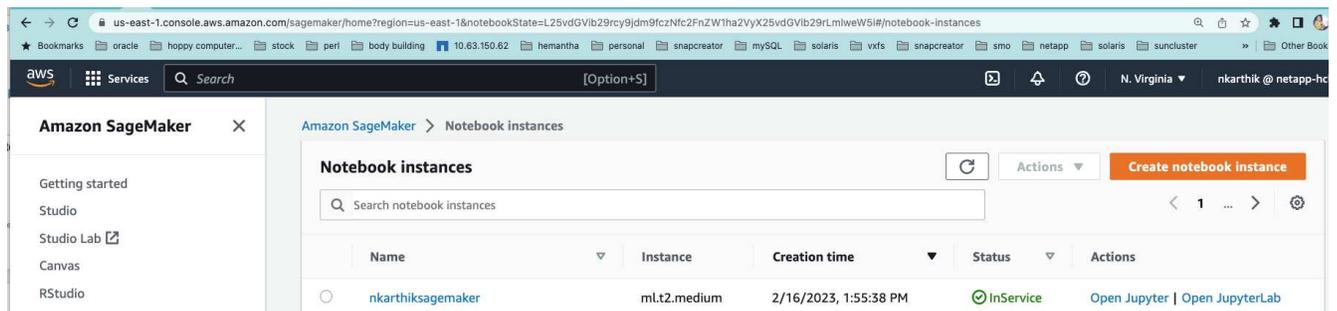
Security Group(s)

[sg-07111a8c16d67c81d](#)

Direct internet access

Enabled: [Learn more](#)

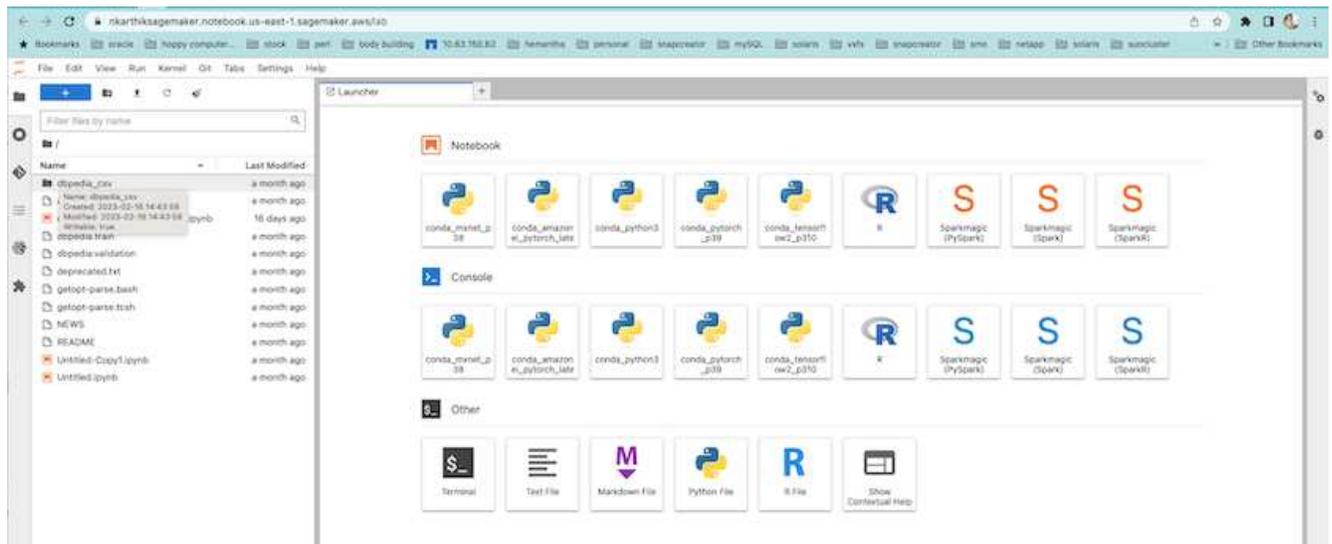
4. Starten Sie das AWS-Notebook.



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

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

5. Öffnen Sie das Jupyter-Labor.



6. Melden Sie sich beim Terminal an und mounten Sie das Cloud Volumes ONTAP -Volume.

```
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. Überprüfen Sie den auf dem Cloud Volumes ONTAP -Volume erstellten Bucket mithilfe der AWS CLI-Befehle.

```
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$
```

Daten für maschinelles Lernen

Bei dieser Validierung haben wir einen Datensatz von DBpedia verwendet, einem Crowdsourcing-Community-Projekt, um strukturierte Inhalte aus den in verschiedenen Wikimedia-Projekten erstellten Informationen zu extrahieren.

1. Laden Sie die Daten vom DBpedia-GitHub-Speicherort herunter und extrahieren Sie sie. Verwenden Sie dasselbe Terminal wie im vorherigen Abschnitt.

```

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. Kopieren Sie die Daten an den Cloud Volumes ONTAP -Speicherort und überprüfen Sie sie mithilfe der AWS CLI aus dem S3-Bucket.

```

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. Führen Sie eine grundlegende Validierung durch, um sicherzustellen, dass die Lese-/Schreibfunktion im S3-Bucket funktioniert.

```

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 /vol1
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 /vol1/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 /vol1/dbpedia_csv
sh-4.2$ ls -ltr /vol1/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$

```

Validieren Sie maschinelles Lernen aus Jupyter Notebooks

Die folgende Validierung ermöglicht das Erstellen, Trainieren und Bereitstellen von Modellen für maschinelles Lernen durch Textklassifizierung unter Verwendung des folgenden SageMaker BlazingText-Beispiels:

1. Installieren Sie die Pakete boto3 und SageMaker.

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

Ausgabe:

```

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,>=1.29.72->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
520df034fbf2d44a1d7ba0681b27ef45aa61
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. Im nächsten Schritt werden die Daten(`dbpedia_csv`) wird aus dem S3-Bucket heruntergeladen `ontapbucket1` zu einer Jupyter Notebook-Instanz, die im maschinellen Lernen verwendet wird.

```

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. Der folgende Code erstellt die Zuordnung von ganzzahligen Indizes zu Klassenbezeichnungen, die zum Abrufen des tatsächlichen Klassennamens während der Inferenz verwendet werden.

```

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()

```

Die Ausgabe listet die Dateien und Ordner in der `ontapbucket1` Bucket, die als Daten für die AWS SageMaker-Maschinenlernvalidierung verwendet werden.

```
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. Starten Sie die Datenvorverarbeitungsphase, um die Trainingsdaten in ein durch Leerzeichen getrenntes, tokenisiertes Textformat vorzuverarbeiten, das vom BlazingText-Algorithmus und der NLTK-Bibliothek verwendet werden kann, um die Eingabesätze aus dem DBPedia-Datensatz zu tokenisieren. Laden Sie den NLTK-Tokenizer und andere Bibliotheken herunter. Der `transform_instance` Die parallele Anwendung auf jede Dateninstanz verwendet das Python-Multiprocessing-Modul.

```
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)

```

Ausgabe:

```

[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. Laden Sie den formatierten Trainingsdatensatz auf S3 hoch, damit er von SageMaker zum Ausführen von Trainingsaufträgen verwendet werden kann. Laden Sie dann mithilfe des Python SDK zwei Dateien in den Bucket hoch und präfixieren Sie den Speicherort.

```

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)

```

Ausgabe:

```

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

```

6. Richten Sie bei S3 einen Ausgabeort ein, an dem das Modellartefakt geladen wird, sodass Artefakte die Ausgabe des Trainingsjobs des Algorithmus sein können. Erstellen Sie ein `sageMaker.estimator.Estimator` Objekt, um den Trainingsjob zu starten.

```

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))

```

Ausgabe:

```

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 framework/algorithm version: latest.
Using SageMaker BlazingText container: 811284229777.dkr.ecr.us-east-
1.amazonsaws.com/blazingtext:1 (us-east-1)

```

7. Definieren Sie den SageMaker `Estimator` mit Ressourcenkonfigurationen und Hyperparametern, um die Textklassifizierung im DBPedia-Datensatz im überwachten Modus auf einer c4.4xlarge-Instanz zu trainieren.

```

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. Bereiten Sie einen Handshake zwischen den Datenkanälen und dem Algorithmus vor. Erstellen Sie dazu die `sagemaker.session.s3_input` Objekte aus den Datenkanälen und speichern Sie sie in einem Wörterbuch, damit der Algorithmus sie verwenden kann.

```

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. Nachdem der Auftrag abgeschlossen ist, wird die Meldung „Auftrag abgeschlossen“ angezeigt. Das trainierte Modell befindet sich im S3-Bucket, der als `output_path` im Schätzer.

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

Ausgabe:

```
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
BlazingText 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/train/dbpedia.train . File size: 35.0693244934082 MB
[02/16/2023 20:39:41 INFO 140279908747072] Processing
/opt/ml/input/data/validation/dbpedia.validation . File size:
21.887572288513184 MB
Read 6M words
Number of words: 149301
Loading validation data from
/opt/ml/input/data/validation/dbpedia.validation
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. Stellen Sie nach Abschluss des Trainings das trainierte Modell als in Echtzeit gehosteten Amazon SageMaker-Endpoint bereit, um Vorhersagen zu treffen.

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

Ausgabe:

```
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. Standardmäßig gibt das Modell eine Vorhersage mit der höchsten Wahrscheinlichkeit zurück. Um die Spitze abzurufen k Vorhersagen, Set k in der Konfigurationsdatei.

```
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. Löschen Sie den Endpunkt, bevor Sie das Notebook schließen.

```
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
```

Abschluss

Basierend auf dieser Validierung können Datenwissenschaftler und Ingenieure über S3-Buckets von NetApp Cloud Volumes ONTAP auf NFS-Daten von AWS SageMaker Jupyter Notebooks zugreifen. Dieser Ansatz ermöglicht den einfachen Zugriff und die gemeinsame Nutzung derselben Daten sowohl von NFS als auch von S3, ohne dass zusätzliche Software erforderlich ist.

Wo Sie weitere Informationen finden

Weitere Informationen zu den in diesem Dokument beschriebenen Informationen finden Sie in den folgenden Dokumenten und/oder auf den folgenden Websites:

- Textklassifizierung mit SageMaker BlazingText
- ONTAP -Versionsunterstützung für S3-Objektspeicher

["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)

Copyright-Informationen

Copyright © 2025 NetApp. Alle Rechte vorbehalten. Gedruckt in den USA. Dieses urheberrechtlich geschützte Dokument darf ohne die vorherige schriftliche Genehmigung des Urheberrechtinhabers in keiner Form und durch keine Mittel – weder grafische noch elektronische oder mechanische, einschließlich Fotokopieren, Aufnehmen oder Speichern in einem elektronischen Abrufsystem – auch nicht in Teilen, vervielfältigt werden.

Software, die von urheberrechtlich geschütztem NetApp Material abgeleitet wird, unterliegt der folgenden Lizenz und dem folgenden Haftungsausschluss:

DIE VORLIEGENDE SOFTWARE WIRD IN DER VORLIEGENDEN FORM VON NETAPP ZUR VERFÜGUNG GESTELLT, D. H. OHNE JEGLICHE EXPLIZITE ODER IMPLIZITE GEWÄHRLEISTUNG, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE STILLSCHWEIGENDE GEWÄHRLEISTUNG DER MARKTGÄNGIGKEIT UND EIGNUNG FÜR EINEN BESTIMMTEN ZWECK, DIE HIERMIT AUSGESCHLOSSEN WERDEN. NETAPP ÜBERNIMMT KEINERLEI HAFTUNG FÜR DIREKTE, INDIREKTE, ZUFÄLLIGE, BESONDERE, BEISPIELHAFT SCHÄDEN ODER FOLGESCHÄDEN (EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE BESCHAFFUNG VON ERSATZWAREN ODER -DIENSTLEISTUNGEN, NUTZUNGS-, DATEN- ODER GEWINNVERLUSTE ODER UNTERBRECHUNG DES GESCHÄFTSBETRIEBS), UNABHÄNGIG DAVON, WIE SIE VERURSACHT WURDEN UND AUF WELCHER HAFTUNGSTHEORIE SIE BERUHEN, OB AUS VERTRAGLICH FESTGELEGTER HAFTUNG, VERSCHULDENSUNABHÄNGIGER HAFTUNG ODER DELIKTSHAFTUNG (EINSCHLIESSLICH FAHRLÄSSIGKEIT ODER AUF ANDEREM WEGE), DIE IN IRGEND EINER WEISE AUS DER NUTZUNG DIESER SOFTWARE RESULTIEREN, SELBST WENN AUF DIE MÖGLICHKEIT DERARTIGER SCHÄDEN HINGEWIESEN WURDE.

NetApp behält sich das Recht vor, die hierin beschriebenen Produkte jederzeit und ohne Vorankündigung zu ändern. NetApp übernimmt keine Verantwortung oder Haftung, die sich aus der Verwendung der hier beschriebenen Produkte ergibt, es sei denn, NetApp hat dem ausdrücklich in schriftlicher Form zugestimmt. Die Verwendung oder der Erwerb dieses Produkts stellt keine Lizenzierung im Rahmen eines Patentrechts, Markenrechts oder eines anderen Rechts an geistigem Eigentum von NetApp dar.

Das in diesem Dokument beschriebene Produkt kann durch ein oder mehrere US-amerikanische Patente, ausländische Patente oder anhängige Patentanmeldungen geschützt sein.

ERLÄUTERUNG ZU „RESTRICTED RIGHTS“: Nutzung, Vervielfältigung oder Offenlegung durch die US-Regierung unterliegt den Einschränkungen gemäß Unterabschnitt (b)(3) der Klausel „Rights in Technical Data – Noncommercial Items“ in DFARS 252.227-7013 (Februar 2014) und FAR 52.227-19 (Dezember 2007).

Die hierin enthaltenen Daten beziehen sich auf ein kommerzielles Produkt und/oder einen kommerziellen Service (wie in FAR 2.101 definiert) und sind Eigentum von NetApp, Inc. Alle technischen Daten und die Computersoftware von NetApp, die unter diesem Vertrag bereitgestellt werden, sind gewerblicher Natur und wurden ausschließlich unter Verwendung privater Mittel entwickelt. Die US-Regierung besitzt eine nicht ausschließliche, nicht übertragbare, nicht unterlizenzierbare, weltweite, limitierte unwiderrufliche Lizenz zur Nutzung der Daten nur in Verbindung mit und zur Unterstützung des Vertrags der US-Regierung, unter dem die Daten bereitgestellt wurden. Sofern in den vorliegenden Bedingungen nicht anders angegeben, dürfen die Daten ohne vorherige schriftliche Genehmigung von NetApp, Inc. nicht verwendet, offengelegt, vervielfältigt, geändert, aufgeführt oder angezeigt werden. Die Lizenzrechte der US-Regierung für das US-Verteidigungsministerium sind auf die in DFARS-Klausel 252.227-7015(b) (Februar 2014) genannten Rechte beschränkt.

Markeninformationen

NETAPP, das NETAPP Logo und die unter <http://www.netapp.com/TM> aufgeführten Marken sind Marken von NetApp, Inc. Andere Firmen und Produktnamen können Marken der jeweiligen Eigentümer sein.