



NetApp 스토리지 플랫폼 전반에 걸쳐 **AI** 파이프라인 간소화

NetApp artificial intelligence solutions

NetApp
September 25, 2026

목차

NetApp 스토리지 플랫폼 전반에 걸쳐 AI 파이프라인 간소화	1
NetApp 스토리지 플랫폼(StorageGRID, ONTAP 및 E-Series) 전반에 걸쳐 AI 파이프라인 간소화	1
1. 요약	1
2. 솔루션 개요, 비즈니스 가치 및 고객 혜택	3
3. 아키텍처 개요	5
4. NetApp, 타사 및 오픈 소스 기술 구성 요소	9
5. 솔루션 설계 및 스토리지 아키텍처 상세 설명	10
6. NetApp XCP를 이용한 데이터 이동성	14
7. 배포 아키텍처	17
8. 구성 참조	22
9. 운영 지침	38
10. 보안 고려 사항, 검증 및 테스트	40
11. NetApp 차별화 및 경쟁력, 부록	44

NetApp 스토리지 플랫폼 전반에 걸쳐 AI 파이프라인 간소화

NetApp 스토리지 플랫폼(StorageGRID, ONTAP 및 E-Series) 전반에 걸쳐 AI 파이프라인 간소화

Karthikeyan Nagalingam, NetApp

이 검증된 아키텍처는 Apache Airflow로 오케스트레이션되는 스토리지 계층 인식 AI 데이터 파이프라인을 제공합니다. 이 솔루션은 NetApp StorageGRID에서 원시 테이블 형식 및 텍스트 데이터를 수집하고, ONTAP NAS 버킷에 결정론적 데이터 세트와 매니페스트를 준비한 다음, NetApp XCP를 사용하여 런타임 구성에 따라 준비된 데이터를 NetApp ONTAP S3 또는 LustreFS로 이동합니다. 그런 다음 scikit-learn 모델을 학습하고 점진적으로 미세 조정하며, 예측을 생성하고, 성공적인 학습 결과를 체크포인트에 저장하고, 모델, 메트릭, 매니페스트 및 예측 증거를 StorageGRID에 아카이빙합니다. 이 아키텍처는 AI 수명 주기 전반에 걸쳐 재현 가능한 실행, 구성 가능한 스토리지 배치, 장애 방지 장치, 체크포인트 기반 복구 및 감사 가능한 데이터 계보를 제공합니다. 이 솔루션은 NetApp 스토리지 계층 전반에 걸쳐 AI 파이프라인을 위한 반복 가능한 배포 및 검증 패턴이 필요한 아키텍트, 데이터 엔지니어, 머신러닝 엔지니어 및 인프라 팀을 위해 설계되었습니다.

목차

1. "요약"
2. "솔루션 개요, 비즈니스 가치 및 고객 혜택"
3. "아키텍처 개요"
4. "NetApp, 타사 및 오픈 소스 기술 구성 요소"
5. "솔루션 설계 및 스토리지 아키텍처 상세 정보"
6. "NetApp XCP를 사용한 데이터 이동성"
7. "배포 아키텍처"
8. "구성 참조"
9. "운영 가이드"
10. "보안 고려 사항, 검증 및 테스트"
11. "NetApp 차별화 및 경쟁력 강화 전략, 부록"

1. 요약

Karthikeyan Nagalingam, NetApp

[[1-1-purpose-of-this-document]]

== 1.1 본 문서의 목적

이 NetApp 검증 아키텍처(NVA)는 원시 데이터를 수집하고, 머신 러닝에 적합하도록 데이터를 준비하고, NetApp XCP를 사용하여 스토리지 계층 간에 고속으로 데이터를 이동하고, 모델을 학습 및 점진적으로 미세 조정하고, 예측을 생성하고, 거버넌스 및 재현성을 위해 결과를 아카이빙하는 프로덕션 수준의 AI 데이터 파이프라인을 문서화합니다. 이

문서는 아키텍트와 인프라 팀이 설계 근거, 배포, 구성 및 운영을 안내하기 위한 것입니다.

[[1-2-audience]]
== 1.2 대상 독자

솔루션 아키텍트, 스토리지/인프라 엔지니어, 데이터 플랫폼 엔지니어, ML 엔지니어 및 AI/ML 워크로드를 위한 NetApp 스토리지 및 데이터 이동성 기술을 평가하는 IT 의사 결정권자.

[[1-3-business-challenge-overview]]
== 1.3 비즈니스 과제 개요

AI 팀은 오브젝트 스토리지에 저장된 데이터를 안정적으로 변환한 후, 일반적인 워크로드에는 탄력적인 S3를, I/O 집약적인 학습에는 고성능 병렬 파일 시스템(LustreFS)을 사용하는 등 학습에 가장 적합한 컴퓨팅-스토리지 계층으로 전달해야 합니다. 관리되는 데이터 이동성 및 오케스트레이션 계층이 없다면, 팀은 재시도, 체크포인트, 스토리지 유연성 및 감사 추적 기능이 부족한 취약하고 단일 목적의 스크립트에 의존하게 됩니다.

기업의 AI 프로그램이 정체되는 이유는 모델이나 컴퓨팅 성능의 한계 때문이 아니라, 기존 스토리지가 AI에 최적화되어 있지 않기 때문입니다. 대규모 데이터셋을 이동하고, 학습 중 GPU에 지속적으로 데이터를 공급하고, 체크포인트 오버헤드를 관리하고, 스토리지 비용을 통제하는 데 드는 노력이 AI 작업 자체보다 더 많습니다. 단일 계층 아키텍처는 성능과 비용 사이의 절충점을 강요하며, 모델과 데이터셋이 커질수록 이러한 절충점은 더욱 심화됩니다.

[[1-4-netapp-solution-summary]]
== 1.4 NetApp 솔루션 요약

이 솔루션은 Apache Airflow DAG(방향 비순환 그래프)를 사용하여 데이터 수집, 준비, 이동, 학습, 미세 조정, 추론 및 아카이빙을 통합적으로 관리합니다. StorageGRID는 원시 데이터 및 장기 아카이빙 오브젝트 계층을 담당합니다. 데이터 준비 과정에서는 실행 스탬프가 찍힌 데이터셋과 매니페스트를 ONTAP NAS 버킷에 기록하고, 이 버킷은 NFS를 통해 XCP 소스로 노출됩니다. NetApp XCP는 준비된 데이터를 런타임에 선택 가능한 학습 대상(NetApp ONTAP S3 또는 LustreFS)으로 이동시키며, 계층 전환 시 코드 변경이 필요하지 않습니다.



이 설계에서 ONTAP NAS 버킷은 주로 XCP 지원 및 유효성 검사 관점에서 사용됩니다. 이는 모빌리티 워크플로우에 일관된 NFS 소스를 제공하기 때문입니다. 실제 운영 환경에서는 동일한 활성 데이터를 고성능 RDMA 지원 경로를 통해 직접 제공할 수도 있으므로, ONTAP NAS 계층은 유일한 런타임 액세스 방식이 아니라 운영상 편리하고 지원 가능한 스테이징 패턴으로 간주해야 합니다.

이 파이프라인은 보다 광범위한 *NetApp AI 스토리지 아키텍처*의 구체적인 구현체입니다. 이 아키텍처는 각 AI 워크로드 단계에 맞춰 성능과 비용을 최적화하도록 설계된 3계층 스토리지 프레임워크입니다.

- **E-Series (LustreFS):** GPU 집약적인 모델 학습 및 체크포인트 작업을 위한 초고속 병렬 스토리지입니다.
- **ONTAP (AFF/AFX):** 학습, 미세 조정, 추론, 벡터/RAG 워크로드 및 일부 데이터 준비 사용 사례를 위한 고성능 액티브 스토리지로, FAS 및 NAS 버킷을 지원합니다.
- **StorageGRID:** 데이터 수집, 거버넌스 및 장기 보존을 위한 확장 가능한 S3 호환 오브젝트 스토리지입니다.

Apache Airflow와 NetApp XCP를 통해 계층 간 데이터 이동이 완전히 자동화되어 핵심 경로에서 수동 작업이 완전히 제거됩니다.

NetApp 스토리지 아키텍처는 기업이 모든 데이터를 단일 중앙 집중식 데이터 레이크에 저장할 필요 없이 분산된 데이터 준비, 모델 학습, 미세 조정, 추론 및 수명 주기 관리를 지원할 수 있도록 합니다. 불필요한 데이터 이동을 줄임으로써 하이브리드 AI 워크플로에 자연스럽게 적합합니다.

[[1-5-why-netapp]]
== 1.5 NetApp을 선택하는 이유

차원	기존 접근 방식	NetApp AI 스토리지 아키텍처
GPU 생산성	스토리지 병목 현상, 유휴 고비용 컴퓨팅	RDMA 기반 GPUDirect Storage를 사용하면 GPU 클러스터를 최대한 활용할 수 있습니다.
데이터 이동성	수동 스크립트는 파이프라인을 지연시킵니다	Airflow + XCP를 통한 자동화된 계층 간 이동
비용 관리	고가 플래시의 모든 데이터	FabricPool 는 사용 빈도가 낮은 데이터를 자동으로 저렴한 오브젝트 스토리지로 이동시킵니다.
멀티 테넌시	공유 네임스페이스는 데이터 유출 위험을 초래합니다.	전용 ONTAP SVM은 엄격한 테넌트 격리를 시행합니다.
워크로드 범위	학습용 스택과 추론용 스택 분리	AI 라이프사이클 전반을 아우르는 단일 통합 아키텍처

[[1-6-the-business-case]]
 == 1.6 비즈니스 케이스

GPU 컴퓨팅은 일반적으로 AI 인프라 예산에서 가장 큰 비중을 차지하는 자본 지출 항목입니다. 클러스터가 데이터를 기다리며 유휴 상태로 있는 매 시간은 직접적이고 측정 가능한 재정적 손실로 이어집니다. NetApp은 적절한 데이터가 적시에 적절한 계층에 자동으로 배치되도록 보장하여 AI 처리량을 높이고 총소유비용(TCO)을 낮추며, 재설계 없이 개념 증명(POC) 단계부터 엔터프라이즈 프로덕션 단계까지 확장 가능한 아키텍처를 제공합니다.

[[1-7-key-benefits-at-a-glance]]
 == 1.7 주요 이점 요약

혜택	설명
통합 오케스트레이션	단일 DAG는 수집 → 아카이브를 아우릅니다.
스토리지 계층 유연성	실행 시 구성을 통해 S3 또는 LustreFS 선택
재교육 비용 절감	~을 통한 점진적 미세 조정 <code>partial_fit</code>
더 빠른 복구	체크포인트 기반 훈련 재개
전체 계보	매니페스트, 유효 구성 로그, 아카이브 레코드
스토리지 잠금 없음	멀티 프로토콜: NFS, S3, Lustre
GPU 사용률	계층형 스토리지를 통해 컴퓨팅 자원을 지속적으로 공급하여 GPU 유휴 시간으로 인한 비용 발생을 방지합니다.
비용 최적화된 유지	FabricPool 콜드 데이터를 오브젝트 스토리지로 자동 계층화합니다

2. 솔루션 개요, 비즈니스 가치 및 고객 혜택

Karthikeyan Nagalingam, NetApp

[[2-1-solution-description]]
 == 2.1 솔루션 설명

파이프라인은 Airflow DAG `example_ai_pipeline_sklearn`로 구현되며, ``ingestion, data_prep, xcp_preflight_source, xcp_copy_prep_to_training, model_training, fine_tuning, inferencing, checkpoint_model_training, route_manual_archive_stage, 및 manual_archive` 작업으로 구성됩니다.

[[2-2-use-case-and-problem-statement]]

== 2.2 사용 사례 및 문제 설명

고객은 StorageGRID에서 가져온 원시 데이터를 사용하여 표 형식 회귀 및 텍스트 분류 모델을 학습시키고 주기적으로 갱신해야 합니다. 파이프라인은 해당 데이터를 ONTAP NAS 버킷에 준비한 다음, 실행별 대상 선택에 따라 학습을 위해 ONTAP S3 또는 LustreFS로 전송하기 위해 NetApp XCP를 사용합니다.

[[2-3-target-customer-profile-and-industry-applicability]]

== 2.3 목표 고객 프로필 및 산업 적용 가능성

금융 서비스(위험/사기 모델), 제조(예측 유지보수), 의료(분류/분류), 소매(수요 예측) 등 NetApp 인프라에서 관리되고 반복 가능한 머신러닝 학습이 필요한 모든 기업에 적합합니다.

[[2-4-solution-scope-and-boundaries]]

== 2.4 솔루션 범위 및 한계

범위 내 항목: 데이터 수집 게이팅, 데이터 준비(Python/Spark), XCP 데이터 이동성(S3/LustreFS), scikit-learn 학습 및 점진적 미세 조정, 추론, 체크포인트 생성, 수동 아카이빙. 범위 외 항목: 실시간 모델 제공 API, 스트리밍 데이터 수집, GPU 기반 딥러닝 프레임워크.

[[2-5-assumptions-and-prerequisites]]

== 2.5 가정 및 전제 조건

- Airflow에서 원시 데이터 액세스 및 아카이빙을 위해 접근 가능한 StorageGRID S3 호환 엔드포인트입니다.
- 준비된 데이터 쓰기를 위해 Airflow에서 접근 가능하고 NFS를 통해 XCP 호스트에 노출되는 ONTAP NAS 버킷.
- NetApp ONTAP S3를 교육 대상으로 선택하면 Airflow 및 XCP에서 ONTAP S3 호환 엔드포인트에 접근할 수 있습니다.
- NetApp XCP가 설치되어 있고 SSH(Airflow 연결 `ssh_default`)를 통해 접속 가능합니다.
- LustreFS가 선택되면 LustreFS 클라이언트가 XCP 호스트에 마운트됩니다.
- Airflow 2.x with `boto3`, `scikit-learn`; 선택 사항: PySpark + Delta/Iceberg 패키지.

[[2-6-business-drivers]]

== 2.6 비즈니스 동인

모델링 시간을 단축하고, 점진적 학습을 통해 인프라 비용을 절감하며, 일회성 스크립팅 위험을 제거합니다.

[[2-7-value-proposition-summary]]

== 2.7 가치 제안 요약

StorageGRID, ONTAP NAS, ONTAP S3, LustreFS 및 XCP를 Airflow 오케스트레이션과 결합하여 스토리지 계층을 인식하고 관찰 가능하며 복원력이 뛰어난 AI 파이프라인을 제공합니다.

[[2-8-stakeholder-benefits]]

== 2.8 이해관계자 혜택

이해관계자	혜택
데이터 엔지니어링	선언적이고 매개변수 기반의 파일 준비; 자동 파일 검색; Spark/Delta/Iceberg 지원
데이터 과학/머신러닝	점진적 미세 조정; 모든 단계에서 일관된 아티팩트 발생원; 재현 가능한 분할
IT 운영	런타임 저장소 선택; 가드/유효 구성 로그; 체크포인트 기반 복구
규정 준수/지배구조	매니페스트 및 보관 기록은 감사 가능한 이력을 제공하며, 보존 단계를 구성할 수 있습니다.

[[2-9-tco-considerations]]

== 2.9 TCO 고려사항

점진적 미세 조정 및 체크포인트 재개 기능을 통해 반복적인 컴퓨팅 비용을 절감할 수 있으며, 스토리지 계층의 유연성을 통해 모든 워크로드에 고성능 스토리지를 과도하게 프로비저닝하는 것을 방지할 수 있습니다.

[[2-10-roi-considerations]]

== 2.10 투자수익률(ROI) 고려사항

모델 업데이트 주기가 빨라지고, 수동 유지 보수가 줄어들며, 장애 발생 시 복구 시간이 단축됩니다.

3. 아키텍처 개요

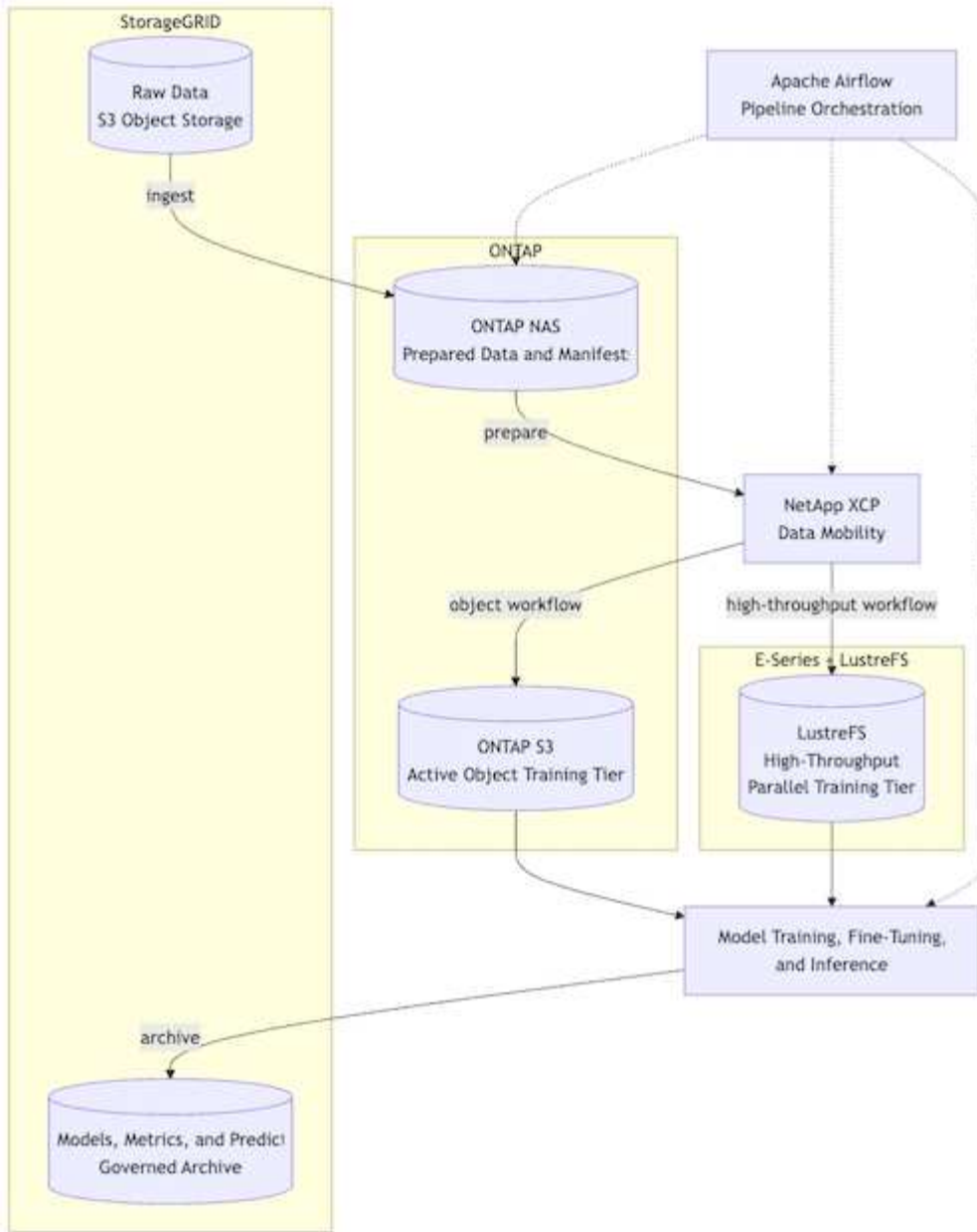
Karthikeyan Nagalingam, NetApp

이 섹션에서는 검증된 AI 파이프라인의 논리적 워크플로, 구성 요소 상호 작용, 물리적 배포 및 데이터 계보를 보여줍니다. 다이어그램을 통해 Apache Airflow가 StorageGRID에서 ONTAP NAS를 거쳐 데이터 이동을 오케스트레이션하고, 선택적으로 NetApp XCP를 사용하여 모델 실행을 위해 ONTAP S3 또는 LustreFS에 데이터를 배치하고, 관리되는 결과를 StorageGRID 아카이브 스토리지에 반환하는 방식을 확인할 수 있습니다.

[[3-1-high-level-three-tier-ai-storage-architecture]]

== 3.1 고수준 3계층 AI 스토리지 아키텍처

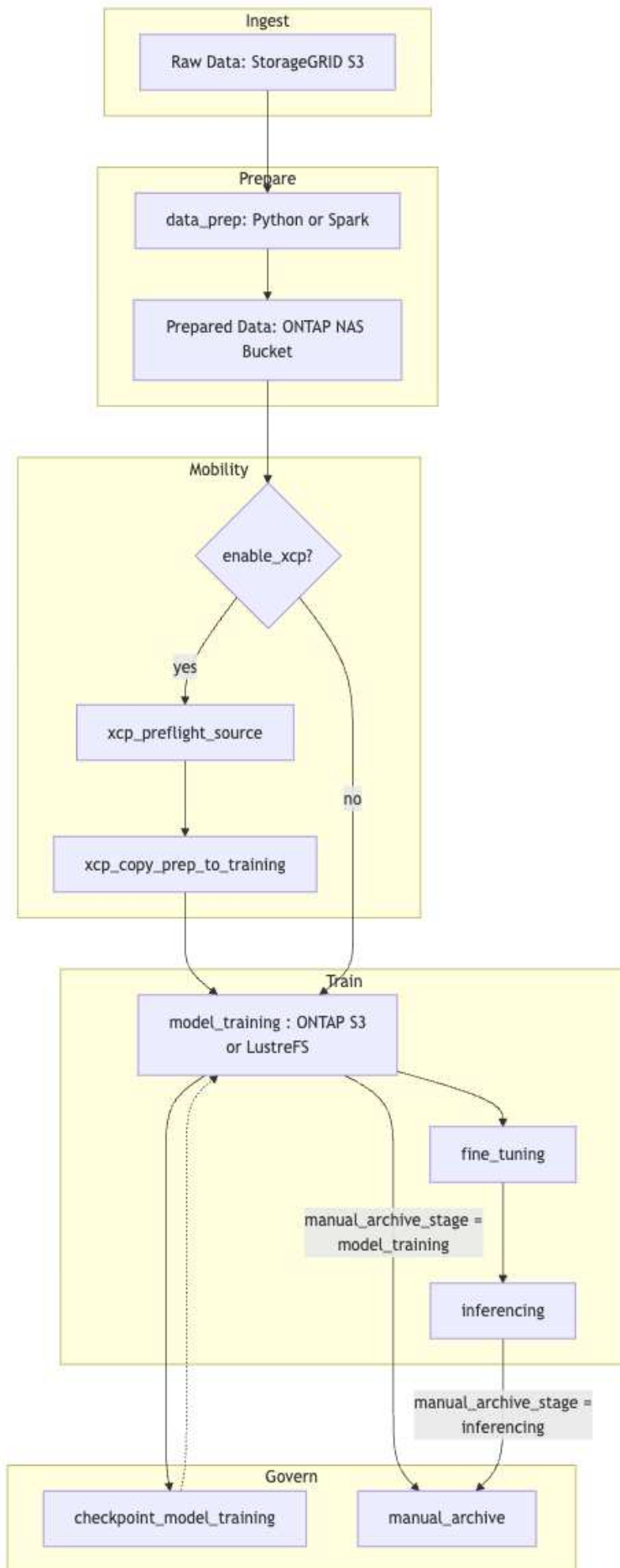
NetApp AI 스토리지 아키텍처는 각 라이프사이클 단계에 가장 적합한 스토리지 계층에 데이터를 배치합니다. StorageGRID는 확장 가능하고 비용 효율적인 원시 데이터 수집 및 관리형 아카이빙에 사용됩니다. ONTAP NAS와 ONTAP S3는 활성 준비, 데이터 이동 스테이징 및 객체 기반 학습 워크플로에 사용됩니다. 높은 처리량의 병렬 학습 I/O가 필요한 경우에는 LustreFS가 탑재된 E-Series가 사용됩니다. Apache Airflow는 라이프사이클 전반을 관리하며, NetApp XCP는 준비된 데이터를 ONTAP NAS 계층에서 선택된 활성 학습 계층으로 이동합니다.



[[3-2-high-level-solution-architecture]]

== 3.2 고수준 솔루션 아키텍처

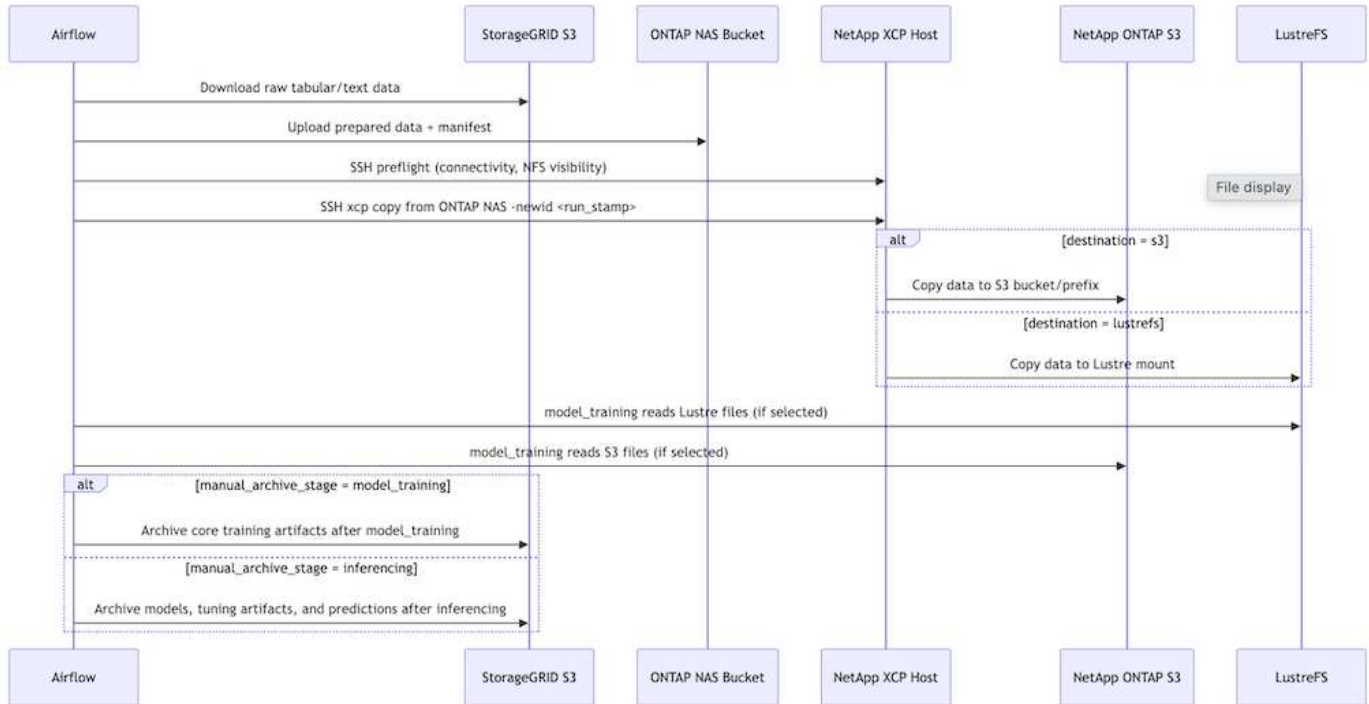
이 워크플로 다이어그램은 Airflow로 제어되는 원시 StorageGRID 데이터부터 준비, 선택적 XCP 이동성, 모델 실행, 체크포인트 생성 및 아카이빙에 이르는 전체 수명 주기를 보여줍니다. `enable_xcp` 준비된 데이터를 학습 전에 ONTAP NAS 버킷에서 ONTAP S3 또는 LustreFS로 복사할지 여부를 결정합니다. 아카이브 분기는 독립적입니다. `manual_archive_stage=model_training` 학습 직후 기준선을 아카이빙하고, `manual_archive_stage=inferencing` 예측이 생성된 후 전체 결과를 아카이빙합니다.



[[3-3-component-interaction-diagram]]

== 3.3 구성 요소 상호 작용 다이어그램

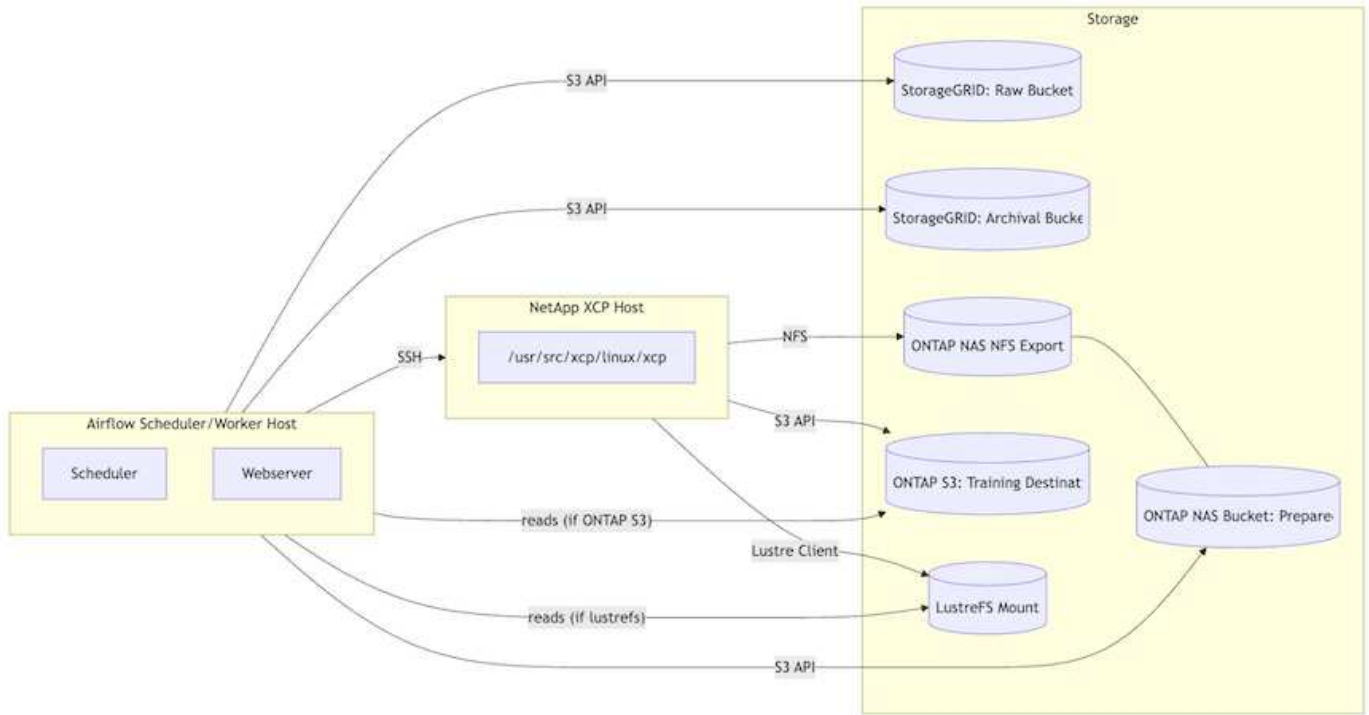
이 시퀀스 다이어그램은 각 데이터 교환을 수행하는 플랫폼 구성 요소를 보여줍니다. Airflow는 StorageGRID에서 원시 객체를 읽고 ONTAP NAS 버킷에 준비된 데이터를 기록한 다음 SSH를 통해 XCP를 호출합니다. XCP는 ONTAP NAS NFS 내보내기를 읽어 구성된 트레이닝 계층에 기록합니다. 마지막 조건 블록은 StorageGRID 아카이브 콘텐츠가 선택된 수동 아카이브 단계에 따라 달라진다는 것을 보여줍니다.



[[3-4-physical-deployment-architecture]]

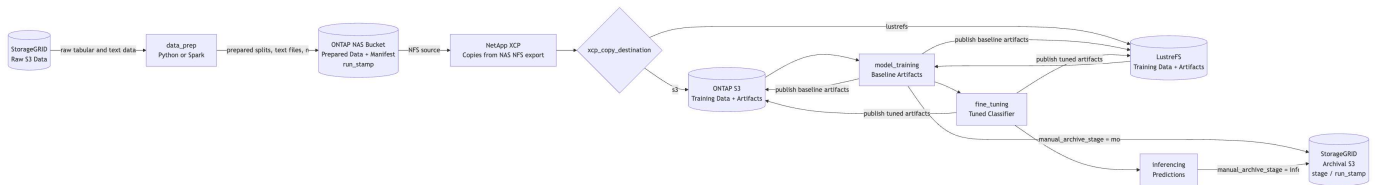
== 3.4 물리적/배포 아키텍처

이 배포 다이어그램은 필요한 호스트, 스토리지 엔드포인트 및 네트워크 인터페이스를 보여줍니다. Airflow 호스트는 StorageGRID, ONTAP NAS 버킷 및 선택된 ONTAP S3 대상에 대한 S3 API를 사용하고 SSH를 사용하여 XCP 호스트를 제어합니다. XCP 호스트는 NFS를 통해 준비된 데이터 소스에 액세스하고 ONTAP S3 또는 LustreFS 마운트에 데이터를 기록합니다. LustreFS가 선택된 경우 Airflow 호스트도 학습 데이터를 읽고 아티팩트를 구체화하기 위해 LustreFS를 마운트해야 합니다.



[[3-5-data-flow-overview]]
 == 3.5 데이터 흐름 개요

이 간결한 흐름도는 데이터 및 아티팩트의 출처를 요약합니다. 각 실행에는 고유한 `run\_stamp`가 할당되어 준비된 데이터, XCP 출력, 모델 아티팩트, 예측 및 아카이브된 객체가 동일한 실행과 연관되어 있음을 추적할 수 있습니다. XCP 대상 선택은 학습 데이터 및 모델 아티팩트 계층을 결정하며, `manual\_archive\_stage`는 StorageGRID가 기본 학습 패키지를 수신할지 전체 추론 패키지를 수신할지를 결정합니다.



4. NetApp, 타사 및 오픈 소스 기술 구성 요소

Karthikeyan Nagalingam, NetApp

요소	솔루션에서의 역할
NetApp StorageGRID	S3 호환 원시 데이터 수집 및 아카이브 오브젝트 스토리지 계층
NetApp ONTAP NAS 버킷 / NFS	준비된 데이터 계층; NFS 소스 내보내기는 데이터 이동 전에 XCP 프리플라이트에 의해 검증됨
NetApp ONTAP S3	XCP 학습 데이터 및 모델 아티팩트의 저장 위치(선택 사항)
NetApp XCP	고성능 데이터 이동 엔진; S3 또는 LusterFS로 복사
NetApp FabricPool	비용 최적화를 위해 ONTAP NAS 준비 데이터 및 아카이브 인접 데이터에 대한 선택적 계층화

요소	솔루션에서의 역할
LustreFS(NetApp 통합)	훈련 계층을 위한 고성능 병렬 파일 시스템
NetApp Console(선택 사항)	ONTAP 및 XCP 리소스에 대한 확장된 모니터링/관찰 가능성
ONTAP 스냅샷/백업	파이프라인과 무관하게 ONTAP NAS 계층에 준비된 데이터에 대한 특정 시점 보호

[[4-1-netapp-xcp-core-value]]
 == 4.1 NetApp XCP — 핵심 가치

XCP는 데이터 위치와 컴퓨팅 위치를 분리하는 모빌리티 계층입니다. 단일 구성 스위치 (`xcp_copy_destination`)를 통해 준비된 데이터와 모델 아티팩트가 NetApp ONTAP S3 또는 LustreFS 고성능 계층에 저장될지 여부가 결정되며, `model_training`, `fine_tuning`, 및 `'inferencing'`는 동일한 대상에서 자동으로 읽기 및 쓰기 작업을 수행합니다.

[[4-2-third-party-and-open-source-technology-components]]
 == 4.2 타사 및 오픈 소스 기술 구성 요소

요소	목적
아파치 에어플로우	DAG 오케스트레이션, 재시도, XCom 태스크 통신, 스케줄링
Apache Spark	<code>'data_prep'</code> 용 선택적 분산 변환 엔진
Delta Lake / Apache Iceberg	준비된 테이블 형식 데이터를 위한 선택적 레이크하우스 테이블 형식
scikit-learn	LinearRegression (표 형식) 및 SGDClassifier (텍스트, 증분 형식)
Airbyte / Apache NiFi	선택적 수집 커넥터가 앞에 있습니다. <code>data_prep</code>
boto3	StorageGRID, ONTAP NAS 버킷 및 ONTAP S3 상호 작용에 사용되는 AWS S3 호환 SDK

5. 솔루션 설계 및 스토리지 아키텍처 상세 설명

Karthikeyan Nagalingam, NetApp

[[5-1-design-principles]]
 == 5.1 설계 원칙

코드 변경보다 구성 우선, 타입이 지정된 계약을 통한 명확한 단계 경계, 일관된 아티팩트 라우팅, 실행 가능한 오류를 통한 신속한 실패 검증, 스토리지 계층 중립성.

[[5-2-stage-design-summary]]
 == 5.2 단계 설계 요약

단계	설계 요약
수집	<code>ingestion_tool s3 direct, none, airbyte</code> 또는 <code>'nifi'</code> 를 선택하고 정규화된 메타데이터를 반환합니다

단계	설계 요약
데이터 준비	원시 StorageGRID 테이블 형식 키를 검색/수락하고, Airbyte/일반 CSV를 정규화하며, 결정론적 학습/검증/추론 분할을 생성하고, 실행 스탬프가 찍힌 준비된 데이터 접두사와 매니페스트를 ONTAP NAS 버킷에 기록합니다.
데이터 이동성(XCP)	ONTAP NAS NFS 내보내기에서 준비된 데이터를 복사하고, xcp_copy_destination XCP 대상 및 학습 읽기 경로를 구동합니다.
모델 학습	로컬 또는 XCP 대상(S3/LustreFS)에서 학습하며, 다양한 형식(CSV/Parquet/JSON) 검색을 지원합니다.
미세 조정	XCP 대상에서 기본 모델을 구체화하고, partial_fit 확장된 클래스 세트로 조정된 모델을 다시 게시합니다.
추론	동일한 XCP 대상에서 조정된 아티팩트를 구체화합니다. 분할/텍스트 범위를 구성할 수 있습니다.
체크포인트/재개	write_stage_checkpoint 완료 기록을 유지하고, _small_resume_guard 후속 실행 시 유효성을 검사하고 재사용합니다.
보관용	model_training 핵심 학습된 모델을 즉시 아카이브하며, inferencing 예측을 기다린 후 전체 결과 세트를 아카이브합니다. 선택적 XCP 입력 포함 및 원시 폴더 정리 기능을 지원합니다.

[[5-3-configuration-driven-philosophy]] == 5.3 구성 중심 철학

비밀 저장소, 엔진 및 동작 선택 사항을 제외한 모든 항목은 dag_run.conf 매개변수입니다. S3와 LustreFS 또는 Python과 Spark 간 전환에는 코드 수정이 필요하지 않으며, Airflow에서 관리하는 비밀 저장소는 필요한 자격 증명을 해결합니다.

[[5-4-storage-architecture-detail]] == 5.4 스토리지 아키텍처 세부 정보

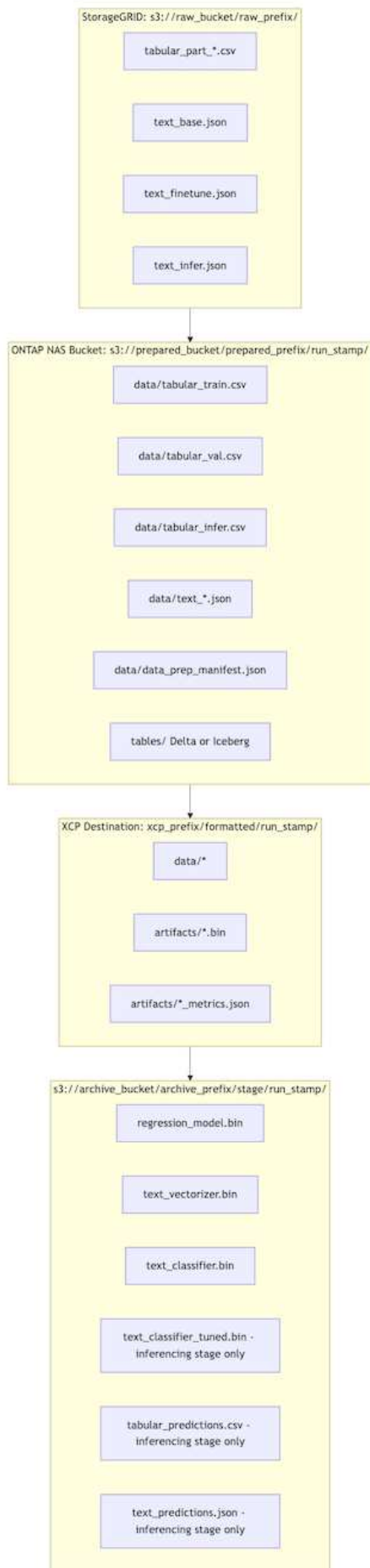
스토리지 아키텍처는 AI 라이프사이클을 목적에 맞게 설계된 계층으로 분리합니다. StorageGRID는 원시 데이터와 아카이브 객체를 저장하고, ONTAP NAS 버킷은 준비되고 실행 스탬프가 찍힌 데이터 세트를 저장하며, ONTAP S3 또는 LustreFS는 선택된 활성 학습 계층을 제공합니다. 이러한 분리를 통해 소스 데이터, 준비된 데이터, 모델 아티팩트 및 아카이브된 증거를 독립적으로 관리할 수 있으며, 공유된 `run\_stamp`을 통해 데이터 계보를 보존합니다.

[[5-4-1-storage-layout-diagrams]] === 5.4.1 스토리지 레이아웃 다이어그램

층	논리적 저장 경로	저장된 내용/아티팩트
StorageGRID 원시 계층	s3://raw_bucket/raw_prefix /	표 형식 CSV 부분 (tabular_part_*.csv), 텍스트 입력 (text_base.json, text_finetune.json, text_infer.json)

층	논리적 저장 경로	저장된 내용/아티팩트
ONTAP NAS Prepared Tier	s3://prepared_bucket/prepared_prefix/run_stamp/	형식화된 분할 (data/tabular_*.csv), 텍스트 파일, data_prep_manifest.json, Lakehouse 테이블 (tables/Delta/Iceberg)
활성 훈련 단계(XCP Dest)	<xcp_prefix>/formatted/run_stamp/	복사된 데이터 분할, 학습된 모델 바이너리 (artifacts/*.bin), 평가 지표 (artifacts/*_metrics.json)
StorageGRID 아카이빙 티어	s3://archive_bucket/archive_prefix/stage/run_stamp/	단계별 기준선 또는 전체 모델 아티팩트, 예측 증거 (tabular_predictions.csv, text_predictions.json)
계층 간 계보 및 흐름	원시 데이터 → 준비됨 → 활성 단계 → 보관됨	공유 `run_stamp`를 통해 모든 스토리지 계층에 걸쳐 엔드투엔드 데이터 이동 및 아티팩트 계보가 연결됩니다.

이 다이어그램은 모든 스토리지 계층에 걸쳐 단일 파이프라인 실행에 대한 논리적 버킷 및 접두사 레이아웃을 보여줍니다.



[[5-4-2-storage-sizing-guidance]]

=== 5.4.2 스토리지 사이징 지침

각 계층의 규모는 역할 및 보존 정책에 따라 독립적으로 설정해야 합니다. ONTAP NAS와 선택한 XCP 대상은 동시 활성 실행을 수용할 수 있어야 하며, StorageGRID 용량은 주로 원시 데이터 및 아카이브 보존에 따라 결정됩니다. 파이프라인 최대 동시 실행을 계획할 때 임시 작업 용량과 향후 확장을 위한 여유 공간을 포함해야 합니다.

층	크기 기준
StorageGRID 원시 버킷	수집 볼륨 × 보존 기간
ONTAP NAS 준비된 데이터 버킷	준비된 데이터셋 크기 × 유지된 실행 횟수
XCP 대상(S3 또는 Lustre)	최대 동시 학습 데이터셋 크기 + 모델 아티팩트 오버헤드
보관용 버킷	보존 정책 × 아카이브 단계 선택 (model_training = 더 작음; inferencing = 더 큼)

[[5-4-3-storage-performance-considerations]]

=== 5.4.3 스토리지 성능 고려 사항

XCP 대상은 `xcp_copy_destination`을 사용하여 실행 시마다 선택되므로 모든 워크로드를 하나의 계층에 강제로 배치하는 대신 워크로드에 맞춰 스토리지 성능을 조정할 수 있습니다. 행 수가 많은 데이터 세트, 동시 읽기 작업이 많은 경우 또는 I/O 집약적인 학습에는 LustreFS를 선택하십시오. 병렬 파일 시스템이 필요하지 않은 비용 효율적이고 탄력적인 액세스 워크로드에는 ONTAP S3를 선택하십시오. 어느 경우든 XCP는 선택한 학습 계층에 맞춰 데이터와 모델 아티팩트를 정렬된 상태로 유지합니다.

6. NetApp XCP를 이용한 데이터 이동성

Karthikeyan Nagalingam, NetApp

NetApp XCP는 ONTAP NAS 준비 데이터 계층과 활성 학습 계층 간의 제어된 전송 경로를 제공합니다. Airflow는 XCP 호스트를 통해 사전 검사를 수행하고 복사 작업을 호출하며, `xcp_copy_destination` 각 실행마다 ONTAP S3 또는 LustreFS를 선택합니다. 모델 학습, 미세 조정 및 추론에서 동일하게 선택된 대상을 사용하여 데이터와 아티팩트를 일관된 계층에 유지합니다.

[[6-1-why-netapp-xcp]]

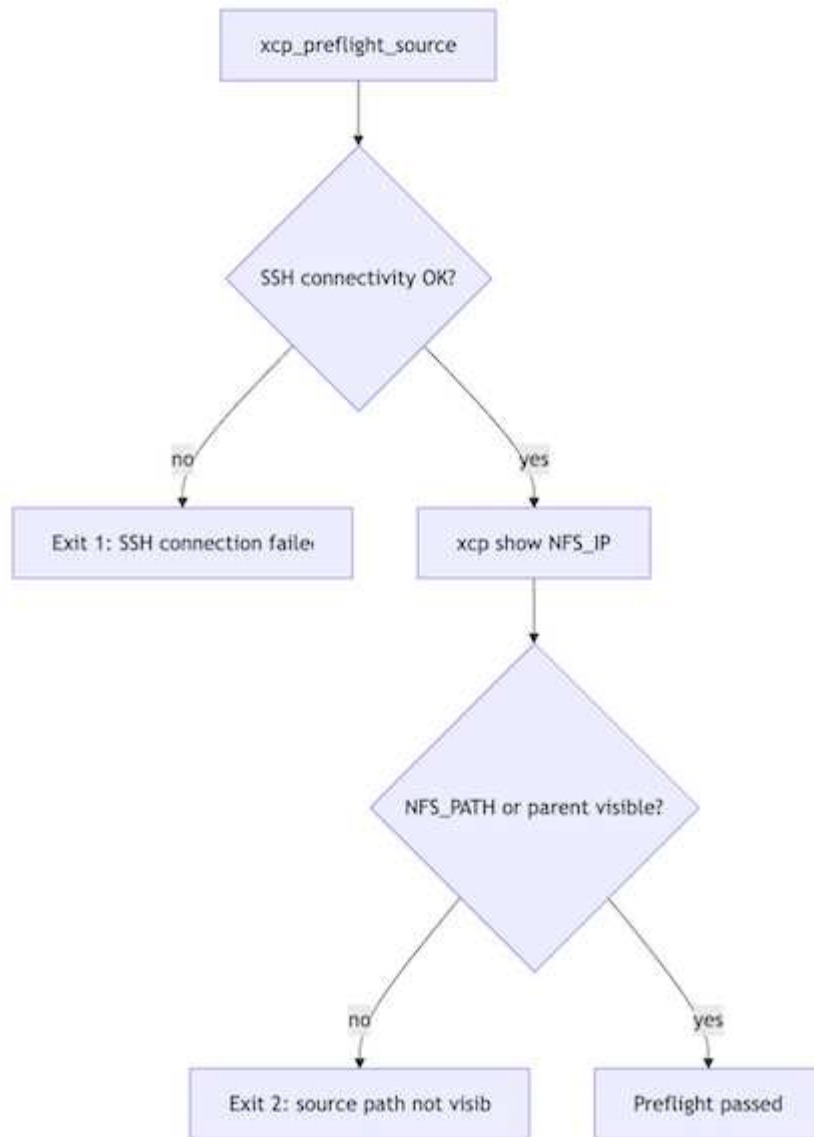
== 6.1 NetApp XCP를 선택하는 이유

XCP는 이 아키텍처에서 사용되는 NFS 소스, ONTAP S3 및 LustreFS 대상을 지원하는 병렬 멀티스레드 복사 엔진입니다. XCP의 `-newid <run_stamp>` 옵션은 각 전송을 파이프라인 실행과 연결하여 준비된 데이터, 학습 입력 및 모델 아티팩트 위치 전반에 걸쳐 추적성을 지원합니다.

[[6-2-xcp-preflight-validation-flow]]

== 6.2 XCP 사전 검증 흐름

사전 점검 작업은 XCP 복사 전에 실행됩니다. 이 작업은 Airflow가 SSH를 통해 XCP 호스트에 연결할 수 있는지, 그리고 구성된 ONTAP NAS NFS 소스 경로가 XCP에서 보이는지 확인합니다. 연결 상태 또는 소스 가시성이 올바르지 않으면 명시적인 종료 조건에 따라 복사가 시작되기 전에 파이프라인이 중지됩니다.



[[6-3-xcp-copy-to-s3-command-template]]

== 6.3 XCP Copy to S3 — 명령 템플릿

이 경로는 `xcp\_copy\_destination=s3`일 때 사용합니다. XCP는 ONTAP NAS NFS 내보내기에서 준비되고 실행 스탬프가 찍힌 데이터를 읽어 구성된 ONTAP S3 버킷 및 접두사에 기록합니다. 명명된 XCP S3 프로파일은 XCP 호스트에 대상 자격 증명을 제공하며, 모델 단계는 `xcp\_s3\_training\_prefix`를 통해 데이터와 아티팩트를 모두 검색합니다.

```

xcp copy -newid <run_stamp> \
  -s3.profile <xcp_s3_profile> \
  -s3.endpoint <xcp_s3_endpoint> \
  <xcp_nfs_source_ip>:<xcp_nfs_source_path> \
  s3://<xcp_s3_bucket>/<xcp_s3_dest_prefix>

```

예시: 준비된 **ONTAP NAS** 데이터셋을 **ONTAP S3** 학습 버킷에 복사합니다.

```
/usr/src/xcp/linux/xcp copy -newid 20260901_120000 \
-s3.profile ontaps3 \
-s3.endpoint http://10.63.150.161 \
10.63.150.159:/prepnasbucket/example_ai_pipeline \
s3://trainingbucket/example_ai_pipeline
```

이 예제에서는 20260901_120000`을 XCP 작업 식별자로 사용합니다. 학습 작업은 구성된 `xcp_s3_training_prefix` 아래에서 복사된 파일을 검색합니다(예: example_ai_pipeline/formatted/20260901_120000/data/ 등).

[[6-4-xcp-copy-to-lustrefs-command-template]]
 == 6.4 XCP Copy to LustreFS — 명령 템플릿

`xcp++\_++copy++\_++destination=lustrefs`일 때 이 경로를 사용합니다. XCP는 마운트된 ONTAP NAS 소스에서 마운트된 LustreFS 대상으로 준비된 데이터를 복사합니다. 두 마운트 모두 XCP 호스트에서 사용 가능해야 하며, LustreFS는 모델 학습, 미세 조정 및 추론을 수행하는 Airflow 워커에도 마운트되어 있어야 합니다.

```
findmnt -T <xcp_lustrefs_source_path> -o FSTYPE,TARGET -n | grep -E
'^nfs[0-9]* ' # source mount type
findmnt -T <xcp_lustrefs_dest_path> -o FSTYPE,TARGET -n | grep -E '^lustre
' # destination mount type
xcp copy -newid <run_stamp> \
file://<xcp_lustrefs_source_path> \
file://<xcp_lustrefs_dest_path>
```

예시: 준비된 **ONTAP NAS** 데이터셋을 **LustreFS** 학습 마운트에 복사합니다.

```
findmnt -T /prepnasbucket/example_ai_pipeline -o FSTYPE,TARGET -n | grep
-E '^nfs[0-9]* '
findmnt -T /mnt/lustre/client -o FSTYPE,TARGET -n | grep -E '^lustre '

/usr/src/xcp/linux/xcp copy -newid 20260901_120000 \
file:///prepnasbucket/example_ai_pipeline \
file:///mnt/lustre/client
```

두 findmnt 명령은 복사가 시작되기 전에 XCP 호스트에서 소스 경로가 NFS를 통해 마운트되고 대상 경로가 LustreFS를 통해 마운트되었는지 확인합니다.

xcp_lustrefs_training_subpath=example_ai_pipeline`를 사용하면 학습은 `/mnt/lustre/client/example_ai_pipeline/formatted/20260901_120000/data/` 아래의 실행 스탬프가 찍힌 데이터를 읽습니다.

[[6-5-xcp-destination-decision-flow]]

== 6.5 XCP 목적지 결정 흐름

이 결정 다이어그램은 런타임 매개변수 하나가 복사 대상과 학습 읽기 경로를 결정하는 방식을 보여줍니다. ONTAP S3 오브젝트 대상의 경우 `s3`를 선택하거나, 병렬 파일 시스템 대상의 경우 `lustrefs`를 선택하십시오. 하위 단계에서는 일치하는 소스를 자동으로 사용하므로 스토리지 위치가 변경되더라도 모델 코드를 수정할 필요가 없습니다.



[[6-6-xcp-failure-handling-and-guardrails]]

== 6.6 XCP 오류 처리 및 안전장치

이러한 안전장치를 통해 구성 및 인프라 오류를 가능한 한 가장 빠른 시점에 관찰할 수 있습니다. 사전 검사 오류는 잘못된 전송을 방지하고, 파일 시스템 유형 검사는 ONTAP NAS 및 LustreFS 마운트를 검증하며, S3 자격 증명 유효성 검사는 ONTAP S3가 선택된 XCP 대상인 경우에만 적용됩니다.

상태	Guard 동작
지원되지 않는 xcp_copy_destination 값	파이프라인은 데이터 이동 전에 값을 거부합니다
LustreFS 마운트가 없습니다	파일 시스템 유형 유효성 검사에 실패했습니다. 데이터 이동 전에 복사가 중단되었습니다.
SSH에 연결할 수 없습니다	사전 검사는 복사 작업 전에 코드 1로 종료됩니다.
NFS 경로가 보이지 않습니다	사전 검사는 복사 작업 전에 코드 2로 종료됩니다.
XCP S3 자격 증명이 누락되었습니다(S3 모드에만 해당)	model_training/artifact sync가 명시적 예외를 발생시킵니다 ValueError

7. 배포 아키텍처

Karthikeyan Nagalingam, NetApp

이 섹션에서는 파이프라인 운영에 필요한 인프라 배치, 소프트웨어 종속성 및 네트워크 연결을 정의합니다. 배포 시 Airflow 오케스트레이션은 XCP 데이터 이동성 호스트와 분리되고, 두 구성 요소는 StorageGRID, ONTAP NAS 및 선택한 ONTAP S3 또는 LustreFS 학습 대상에 연결됩니다.

[[7-1-reference-infrastructure-requirements]]

== 7.1 참조 인프라 요구 사항

다음 구성 요소는 최소 기능 배포 공간을 구성합니다. 선택한 준비 엔진과 동시 작업 부하에 맞게 Airflow 호스트의 크기를 조정하고, XCP 호스트는 불필요한 네트워크 홉 없이 ONTAP NAS NFS 내보내기 및 선택한 대상에 액세스할 수 있는 위치에 배치하십시오. LustreFS 모드를 사용하려면 XCP 호스트와 Airflow 워커 모두에 호환되는 마운트가 필요합니다.

요소	요구 사항
Airflow 호스트	Spark 또는 Python 변환 부하에 맞춰 CPU/메모리 크기 조정
XCP 호스트	NFS/Lustre 및 NetApp ONTAP S3 엔드포인트에 대한 네트워크 액세스
LustreFS 클라이언트	대상으로 선택된 경우 XCP 호스트 및 Airflow 호스트에 마운트됩니다.

[[7-1-1-reference-validation-environment]]

== 7.1.1 참조 검증 환경

다음 환경은 대표적인 단일 노드 검증 프로파일로 사용되었습니다. 이는 기능 및 성능 평가를 위한 출발점일 뿐이며, 실제 운영 환경에 적합한 규모를 권장하는 것은 아닙니다. 고객은 데이터셋 규모, 동시 접속량, 데이터 보존 기간, 복구 기간 및 서비스 수준 요구 사항에 따라 컴퓨팅, 네트워크, 스토리지 용량 및 데이터 보호 수준을 결정해야 합니다.

계층	레퍼런스 하드웨어/소프트웨어
컴퓨팅 및 네트워크	256GB RAM, 64개 CPU 코어, 10GbE 연결을 갖춘 서버 1대
ONTAP 액티브 스토리지	One NetApp A800 시스템 1대 (1.8TB SSD 48개 장착)
오브젝트 스토리지	NetApp StorageGRID SG5864 어플라이언스 1대
고처리량 학습 스토리지	LustreFS가 탑재된 NetApp E2812 시스템 1대
플랫폼 소프트웨어	Kubernetes, Apache Airflow, Apache Airbyte, 단일 노드 Apache Spark, NetApp XCP 및 LustreFS

[[7-2-software-version-matrix]]

== 7.2 소프트웨어 버전 매트릭스

이 매트릭스는 검증된 파이프라인에서 사용되는 런타임 소프트웨어를 식별합니다. Airflow 및 Python 환경은 배포된 공급자 및 패키지와 호환되도록 유지해야 합니다. Spark, Delta Lake 및 Iceberg는 선택 사항이며 `data\_prep` Spark 변환 경로 또는 해당 테이블 형식을 사용하는 경우에만 필요합니다.

소프트웨어	버전/참고 사항
아파치 에어플로우	2.x
파이썬	3.11
boto3	최신 안정 버전
scikit-learn	최신 안정 버전
PySpark(선택 사항)	Delta 3.2.0 / Iceberg 1.5.2 런타임 패키지와 호환됩니다.
NetApp XCP	원격 호스트에 설치됨(예: /usr/src/xcp/linux/xcp

[[7-3-network-requirements]]

== 7.3 네트워크 요구 사항

이러한 네트워크 흐름은 라우팅, 방화벽 및 이름 확인 정책에서 허용되어야 합니다. 프로덕션 환경의 모든 S3 호환 엔드포인트에는 HTTPS 사용을 권장합니다. 엔드포인트가 기본 HTTPS 포트를 사용하지 않는 경우 구성된 사용자 지정

포트를 사용하십시오. 배포된 LustreFS 구현에 따라 Lustre 클라이언트 연결을 확인하십시오.

흐름	프로토콜/포트
Airflow → ONTAP S3	HTTPS/HTTP (443/80 또는 사용자 지정)
Airflow → XCP 호스트	SSH(22)
XCP 호스트 → NFS 소스	NFS(2049)
XCP 호스트 → LustreFS	Lustre 클라이언트 포트
XCP 호스트 → ONTAP S3 (S3 모드)	HTTPS/HTTP

[[7-4-deployment-notes]]

== 7.4 배포 참고 사항

섹션/주제	지침/운영 실무
SSH 연결 설정	포트 22를 통해 XCP 호스트를 대상으로 하는 Airflow 연결 <code>ssh_default`</code> 을 제한된 SSH 키 권한으로 구성합니다(`600).
서비스 수명주기 관리	<code>`tools/airflow_ctl.sh`</code> 를 사용하여 배포 및 유지 관리 중에 로컬 Airflow 스케줄러와 웹 서버의 수명 주기를 관리합니다.
자격 증명 및 비밀 정보 저장	자격 증명, API 토큰 및 액세스 키는 인라인으로 입력하는 대신 Airflow Connections, Variables 또는 시크릿 백엔드에 저장하십시오 <code>dag_run.conf</code> .
인프라 및 작업 관찰 가능성	스케줄러 하트비트와 작업 실행을 독립적으로 모니터링하여 오케스트레이션 가용성 문제를 DAG 장애와 구분합니다.

[[7-5-installation-and-prerequisites]]

== 7.5 설치 및 필수 조건

이 하위 섹션에서는 새로운 환경에서 NetApp AI 검증 파이프라인을 구축, 구성 및 실행하는 데 필요한 기본 설치 단계를 정의합니다. 이 섹션은 DAG를 실행하기 전에 Airflow 워크스페이스를 설정하는 운영자, 아키텍트 및 엔지니어를 대상으로 합니다.

[[7-5-1-prerequisites]]

=== 7.5.1 필수 조건

솔루션을 설치하기 전에 대상 호스트가 다음 최소 요구 사항을 충족하는지 확인하십시오.

- Linux 운영 체제, 가급적 Ubuntu 22.04/24.04 또는 RHEL 8+.
- Python 3.11(`pip`, `venv` 및 빌드 툴링 포함)을 사용할 수 있습니다.
- 소스 코드 검색 및 버전 관리를 위해 호스트에 Git이 설치되어 있습니다.
- Apache Airflow 2.x가 전용 Python 가상 환경에 배포됩니다.
- Airflow 호스트에서 NetApp XCP 호스트로의 SSH 접속.
- Airflow 호스트에서 StorageGRID, ONTAP NAS 및 선택한 ONTAP S3 또는 LustreFS 대상으로 네트워크

연결.

- 원시 데이터, 준비된 데이터 스테이징 및 아카이브 저장소를 위한 S3 호환 엔드포인트 액세스.
- 선택 사항: Java 런타임 및 Spark 3.x(spark 준비 경로를 사용하는 경우).
- 선택 사항: table_format=delta 또는 table_format=iceberg를 선택할 경우 Delta Lake 및 Iceberg 런타임 패키지.
- 선택 사항: Lustre 클라이언트 마운트(xcp_copy_destination=lustrefs 사용 시).

[[7-5-2-software-package-acquisition-and-repository-access]]

=== 7.5.2 소프트웨어 패키지 획득 및 저장소 접근

이 솔루션에 필요한 소프트웨어 패키지, 자동화 DAG, 배포 스크립트 및 검증 아티팩트를 얻으려면 NetApp AI 및 데이터 모빌리티 엔지니어링 팀(ng-data-mobility-in-ai-pipeline@netapp.com)으로 문의하십시오.

접근 권한이 부여되면 GitHub에서 대상 작업 디렉터리로 프로젝트 소스를 다운로드하거나 복제하십시오.

```
# Example GitHub clone
mkdir -p /opt/netapp-ai
cd /opt/netapp-ai

git clone https://github.com/<your-org>/<your-repo>.git
cd <your-repo>

git checkout main
ls -la
```

이미 /opt/netapp-ai/<your-repo> 아래에 저장소를 복제했다면, 다른 복사본을 다운로드하는 대신 해당 작업 디렉터리에서 계속 진행하세요.

[[7-5-3-create-the-python-environment]]

=== 7.5.3 Python 환경 생성

전용 가상 환경을 만들고 Airflow 및 파이프라인에 필요한 기본 종속성을 설치합니다.

```
export REPO_ROOT=/opt/netapp-ai/<your-repo>
cd "$REPO_ROOT"
python3 -m venv "$REPO_ROOT/.venv"
source "$REPO_ROOT/.venv/bin/activate"

python -m pip install --upgrade pip setuptools wheel
pip install "apache-airflow>=2.8,<3.0" boto3 scikit-learn
```

Spark 기반 준비 및 Lakehouse 테이블 형식을 사용하려면 필요에 따라 선택적 패키지를 설치하십시오.

```
pip install pyspark==3.5.*
```

```
pip install delta-spark==3.2.0
pip install apache-iceberg==1.5.2
```

사용 중인 Spark 런타임 및 클러스터 토폴로지에서 지원하는 정확한 패키지 버전을 사용하십시오. 호환되지 않는 Spark, Delta 및 Iceberg 조합을 혼용하지 마십시오.

[[7-5-4-configure-airflow-and-the-repository]]

=== 7.5.4 Airflow 및 저장소 구성

복제된 저장소 루트에서 Airflow 메타데이터 데이터베이스를 초기화하고 DAG 파일이 Airflow에서 보이는지 확인합니다.

```
export AIRFLOW_HOME="$REPO_ROOT/.airflow"
export AIRFLOW__CORE__DAGS_FOLDER="$REPO_ROOT"
export AIRFLOW__CORE__LOAD_EXAMPLES=False

airflow db init
airflow dags list | grep -E
"example_ai_pipeline|example_ai_pipeline_sklearn"
```

워크스페이스에는 검증된 파이프라인에 필요한 Airflow 구성 파일과 DAG가 포함되어 있습니다. 저장소에 로컬 라이프사이클 관리를 위한 헬퍼 스크립트가 있는 경우, Airflow를 수동으로 실행하는 대신 해당 스크립트를 사용하여 스케줄러와 웹 서버를 시작하십시오.

```
# Start the Airflow scheduler and webserver with the deployment's service
manager.
# or, if using the standard commands:
# airflow scheduler
# airflow webserver
```

[[7-5-5-required-connectivity-and-secret-configuration]]

=== 7.5.5 필수 연결 및 비밀 구성

파이프라인을 실행하기 전에 Airflow 호스트에서 다음 리소스에 접근할 수 있는지 확인하십시오.

- StorageGRID 원시 데이터 S3 엔드포인트.
- ONTAP NAS 준비된 데이터 버킷 엔드포인트.
- ONTAP S3 대상 엔드포인트(`xcp_copy_destination=s3`).
- LustreFS 마운트 시점 `xcp_copy_destination=lustrefs`.
- XCP host over SSH using `ssh_default` 또는 구성된 SSH 연결을 사용합니다.

자격 증명을 셸 기록이나 DAG 코드에 직접 삽입하는 대신 Airflow 연결, 변수 또는 시크릿 백엔드에 저장하십시오. 파이프라인 실행을 시작하기 전에 필요한 액세스 키, 엔드포인트 및 프로필 매핑을 설정하십시오.

[[7-5-6-verify-the-installation]]

=== 7.5.6 설치 확인

Airflow 스케줄러와 DAG가 실행되고 예제 파이프라인을 구성 오류 없이 나열하고 실행할 수 있으면 설치가 성공적으로 완료된 것입니다.

```
airflow dags list
airflow tasks list example_ai_pipeline_sklearn
```

설치가 올바르게 완료되면 DAG가 Airflow에 존재하게 되며, 저장소 루트에 있는 트리거 스크립트를 사용하여 `CONF\_JSON` 기반 실행을 위한 준비가 됩니다.

```
./trigger_and_wait_ai_pipeline_sklearn.sh
```

이 시점에서 환경은 8절의 구성 예제와 운영 가이드에 설명된 배포 시나리오를 실행할 준비가 되었습니다.

8. 구성 참조

Karthikeyan Nagalingam, NetApp

구성 참조에는 데이터 수집, 데이터 준비, XCP 이동성, 학습, 체크포인트 및 아카이빙을 제어하는 런타임 매개변수가 설명되어 있습니다. ONTAP NAS 버킷은 주로 지원 및 이식성을 위해 준비된 데이터와 XCP 스테이징 계층으로 사용됩니다. 실제 운영 환경에서는 워크로드 및 지연 시간 요구 사항에 더 적합한 경우 동일한 준비된 데이터를 RDMA 지원 직접 액세스 경로를 통해 사용할 수도 있습니다.

모든 런타임 동작은 `dag\_run.conf`를 통해 제공되므로 동일한 DAG가 서로 다른 소스 시스템, 변환 엔진, 스토리지 대상, 복구 정책 및 아카이브 범위를 지원할 수 있습니다. 선택한 워크플로에 필요한 파라미터 그룹만 구성하고, 자격 증명은 Airflow Connections 또는 비밀 백엔드에 보관하며, 예제는 자격 증명 값이 아닌 비프로덕션 템플릿으로 사용하십시오.

[[8-1-ingestion-parameters]]

== 8.1 수집 매개변수

이 설정을 사용하여 업스트림 데이터 수집 메커니즘을 선택하고 구성하십시오. `s3\_direct`는 데이터 준비 그룹에 구성된 StorageGRID 원시 데이터 객체를 사용합니다. Airbyte 및 NiFi 설정은 해당 수집 툴이 선택된 경우에만 필요합니다.

매개변수	기본값	필수	설명	예
ingestion_tool	s3_direct	아니요	수집 모드를 선택합니다	"s3_direct"
airbyte_api_url	None	Airbyte 전용	Airbyte API 엔드포인트	"http://airbyte:8000"
airbyte_connection_id	None	Airbyte 전용	Airbyte 연결 ID	"connection-123"
airbyte_api_token	None	선택 사항	Bearer 토큰	"<TOKEN>"

매개변수	기본값	필수	설명	예
airbyte_timeout_secs	30	아니요	Airbyte 타임아웃	60
nifi_api_url	None	NiFi 전용	NiFi API 엔드포인트	"http://nifi:8080"
nifi_process_group_id	None	NiFi 전용	NiFi 프로세스 그룹	"abc123"
nifi_timeout_secs	30	아니요	NiFi 타임아웃	60

[[8-2-data-preparation-parameters]]

== 8.2 데이터 준비 매개변수

이 설정은 원시 데이터 액세스, 준비된 데이터 출력, 입력 검색 및 변환 동작을 제어합니다. `s3_raw` * 매개변수는 StorageGRID를 참조하고, 코드 호환 `s3_formatted` * 매개변수는 준비되고 실행 스탬프가 찍힌 데이터와 매니페스트가 기록되는 ONTAP NAS 버킷을 식별합니다.

Raw StorageGRID / prepared-data ONTAP NAS 버킷:

StorageGRID와 ONTAP NAS 버킷이 서로 다른 S3 호환 서비스 또는 액세스 정책을 사용하는 경우 별도의 엔드포인트와 자격 증명을 구성하십시오. 일반적인 대체 키는 보다 구체적인 원시 데이터 또는 준비된 데이터 설정이 없는 경우에만 적용됩니다.

매개변수	기본값	설명	예
s3_raw_bucket	ai-raw-data	StorageGRID 원시 입력 버킷	"bucket1"
s3_raw_prefix	example_ai_pipeline/raw	StorageGRID 원시 데이터 접두사	"example_ai_pipeline/raw"
s3_raw_endpoint_url	폴백 체인	StorageGRID S3 엔드포인트	"http://10.63.150.62:10444"
s3_raw_access_key_id	폴백 체인	원시 액세스 키	"<KEY>"
s3_raw_secret_access_key	폴백 체인	원시 비밀 키	"<SECRET>"
s3_raw_region	aws_region	원시 영역	"us-east-1"
s3_formatted_bucket	ai-formatted-data	ONTAP NAS 준비 데이터 출력 버킷	"prepnasbucket"
s3_formatted_prefix	example_ai_pipeline/formatted	ONTAP NAS 준비 데이터 접두사	"example_ai_pipeline/formatted"
s3_formatted_endpoint_url	폴백 체인	ONTAP NAS 버킷 S3 엔드포인트	"http://10.63.150.159"
s3_formatted_access_key_id	폴백 체인	ONTAP NAS 버킷 액세스 키	"<KEY>"

매개변수	기본값	설명	예
s3_formatted_secret_access_key	폴백 체인	ONTAP NAS 버킷 비밀 키	"<SECRET>"

일반적인 대체값: s3_endpoint_url, s3_access_key_id, s3_secret_access_key, s3_session_token, s3_verify, aws_region, aws_access_key_id, aws_secret_access_key, aws_session_token.

입력 선택:

알려진 CSV 입력을 처리하려면 명시적인 객체 키 설정을 사용하십시오. 그렇지 않으면 작업이 구성된 StorageGRID 원시 접두사 아래에서 적합한 테이블 형식 객체를 검색합니다.

매개변수	기본값	설명	예
s3_tabular_object_keys	자동 검색	CSV 키의 명시적 목록	["raw/demo/a.csv"]
s3_tabular_object_key	None	단일 명시적 CSV 키	"raw/demo/a.csv"

변환:

경량 로컬 준비를 위한 Python 경로 또는 분산 준비를 위한 Spark 경로를 선택하십시오. 샘플링 및 시드는 생성된 학습, 검증 및 추론 분할에 재현 가능하게 적용됩니다.

매개변수	기본값	설명	예
transformation_engine	python	python 또는 spark	"spark"
spark_required	false	폴백 대신 실패	true
table_format	none	none/delta/iceberg	"delta"
table_format_required	false	형식을 사용할 수 없는 경우 실패	true
sample_count	모든 행	행 제한 (≤ 0 = 모두)	14400
seed	7	재현 가능한 셔플 시드	11

Spark:

이 설정은 `transformation\_engine=spark`에만 적용됩니다. Spark 프로세스 배치, 리소스 할당, 시간 제한 및 선택적 Delta Lake 또는 Iceberg 런타임 종속성을 제어합니다.

매개변수	기본값	설명	예
spark_submit_bin	spark-submit	Spark 바이너리	"/opt/spark/bin/spark-submit"
spark_master	local[*]	Spark 마스터	"local[*]"

매개변수	기본값	설명	예
spark_driver_memory	4g	드라이버 메모리	"8g"
spark_executor_memory	4g	실행기 메모리	"8g"
spark_timeout_secs	900/7200	타임아웃 (0 = 비활성화)	0
spark_delta_packages	Delta 3.2.0	델타 런타임 패키지	"io.delta:delta-spark_2.12:3.2.0"
spark_iceberg_packages	Iceberg 1.5.2	Iceberg 런타임 패키지	"org.apache.iceberg:iceberg-spark-runtime-3.5_2.12:1.5.2"
spark_iceberg_catalog	local	Iceberg 카탈로그 이름	"local"

[[8-3-xcp-and-training-destination-parameters]]

== 8.3 XCP 및 교육 대상 매개변수

이 그룹은 `enable\_xcp=true`를 사용하여 ONTAP NAS NFS 소스를 검증하고, XCP 호스트를 호출하고, 활성 교육 계층을 선택할 때 사용합니다. `xcp\_copy\_destination`는 제어 스위치 역할을 하며, 아래의 ONTAP S3 관련 설정 또는 LustreFS 관련 설정을 선택하고, 다운스트림 모델 단계에서는 동일하게 선택된 계층을 사용합니다.

매개변수	기본값	필수	설명	예
enable_xcp	false	네, XCP의 경우 그렇습니다.	XCP 분기 활성화	true
xcp_copy_destination	s3	아니요	s3 또는 lustrefs	"lustrefs"
xcp_nfs_source_ip	10.63.150.159	XCP preflight	NFS 서버 IP	"10.63.150.159"
xcp_nfs_source_path	/prepnasbucket/example_ai_pipeline	XCP preflight	NFS 내보내기 경로	"/prepnasbucket/example_ai_pipeline"
ssh_user	root	아니요	XCP 호스트 SSH 사용자	"root"
ssh_host	10.63.150.178	아니요	XCP 호스트 주소	"10.63.150.178"

S3 대상(필수 조건 **xcp_copy_destination=s3**):

이 설정은 ONTAP S3 교육 경로에만 제공하십시오. 프로필 또는 직접 자격 증명을 사용하면 XCP와 모델 아티팩트 동기화 로직이 선택한 ONTAP S3 버킷 및 접두사에 액세스할 수 있습니다.

매개변수	기본값	설명	예
xcp_s3_endpoint	기본	XCP S3 엔드포인트	"http://10.63.150.161"
xcp_s3_bucket	None	XCP 대상 버킷	"trainingbucket"
xcp_s3_dest_prefix	example_ai_pipeline	목적지 접두사	"example_ai_pipeline"
xcp_s3_training_prefix	xcp_s3_dest_prefix	교육 경로 검색 접두사	"example_ai_pipeline"
xcp_s3_profile	ontaps3	명명된 자격 증명 프로필	"ontaps3"
xcp_s3_profiles	None	프로필 맵	{"ontaps3": {...}}
xcp_s3_access_key_id	None	직접 키 대체	"<KEY>"
xcp_s3_secret_access_key	None	직접 비밀 폴백	"<SECRET>"
xcp_s3_region	aws_region	XCP S3 지역	"us-east-1"

LustreFS 대상 (필수 조건 `xcp_copy_destination=lustrefs`):

이 설정은 LustreFS 교육 경로에만 제공하십시오. 소스 및 대상 경로는 XCP 호스트에 마운트되어 접근 가능해야 하며, 대상 경로는 모델 단계를 실행하는 Airflow 워커에서도 접근 가능해야 합니다.

매개변수	기본값	설명	예
xcp_lustrefs_source_path	/prepnasbucket/example_ai_pipeline	XCP 복사 소스	"/prepnasbucket/example_ai_pipeline"
xcp_lustrefs_dest_path	/mnt/lustre/client	LustreFS 마운트 대상	"/mnt/lustre/client"
xcp_lustrefs_training_subpath	xcp_s3_dest_prefix	마운트 아래의 교육 하위 디렉토리	"example_ai_pipeline"
xcp_lustrefs_newid	`data_prep`런 스탬프	XCP 작업 식별자	"20260901_120000"

텍스트 소스 동작:

이 옵션은 XCP 이동성 이후 세 가지 텍스트 데이터 세트를 찾는 방법을 제어합니다. 명시적 키는 검색을 재정의하며, 로컬 대체 기능을 비활성화하면 텍스트 입력이 선택한 XCP 대상에서 제공되어야 합니다.

매개변수	기본값	설명	예
xcp_text_allow_local_fallback	true	로컬 텍스트 대체 허용	false
xcp_text_base_key	자동	명시적인 기본 텍스트 위치	".../text_base.json"

매개변수	기본값	설명	예
xcp_text_finetune_key	자동	명시적 미세 조정 텍스트 위치	"... /text_finetune.json"
xcp_text_infer_key	자동	명시적 추론 - 텍스트 위치	"... /text_infer.json"

[[8-4-model-training-parameters]]

== 8.4 모델 학습 매개변수

이 설정은 로컬 또는 XCP를 통해 제공되는 소스를 선택하고, 학습 용량을 제한하며, 재현 가능한 데이터 순서를 유지합니다. XCP가 활성화되면 학습은 선택된 ONTAP S3 또는 LustreFS 계층에서 데이터를 읽고 기준선 결과물을 동일한 대상으로 다시 게시합니다.

매개변수	기본값	설명	예
enable_xcp	false	XCP 또는 로컬 소스를 선택합니다	true
xcp_copy_destination	s3	소스 티어를 선택합니다	"lustrefs"
sample_count	모든 행	교육 샘플 제한	14400
seed	7	재현 가능한 셔플	11

생성 항목: regression_model.bin, text_vectorizer.bin, text_classifier.bin, train_metrics.json.

[[8-5-fine-tuning-parameters]]

== 8.5 미세 조정 매개변수

미세 조정은 선택한 XCP 대상에서 기준 텍스트 아티팩트를 구체화하고, 미세 조정 데이터 세트에 증분 학습을 적용한 다음, 조정된 분류기와 해당 지표를 다시 게시합니다. 아티팩트의 출처를 유지하려면 XCP 설정을 모델 학습과 일관되게 유지해야 합니다.

매개변수	기본값	설명	예
enable_xcp	false	아티팩트 구체화/게시를 활성화합니다.	true
xcp_copy_destination	s3	아티팩트 소스/대상을 선택합니다.	"lustrefs"

소비: text_vectorizer.bin, text_classifier.bin, text_finetune.json. 생산: text_classifier_tuned.bin, fine_tune_metrics.json.

[[8-6-inference-parameters]]

== 8.6 추론 매개변수

이 플래그는 선택한 학습 계층에서 튜닝된 모델이 생성된 후의 스코어링 범위를 제어합니다. 검증 또는 비즈니스 요구 사항에 따라 기본 추론 분할 및 추론 텍스트 입력 범위를 넘어서는 출력이 필요한 경우 두 옵션 중 하나를 활성화하십시오.

매개변수	기본값	설명	예
infer_all_splits	false	점수 학습/검증/추론 분할	true
infer_all_text	false	점수 기반/세부 조정/텍스트 추론	true

소비: regression_model.bin, text_vectorizer.bin, text_classifier_tuned.bin. 생성: tabular_predictions.csv, text_predictions.json.

[[8-7-checkpoint-parameters]]

== 8.7 체크포인트 매개변수

체크포인트 기능은 `model\_training`으로 범위가 지정됩니다. 이러한 설정은 학습 완료 기록 생성 여부, 저장 위치, 업로드 실패 처리 방식, 그리고 후속 실행 시 유효한 학습 완료 결과를 검증하거나 재사용할지 여부를 결정합니다.

매개변수	기본값	설명	예
checkpoint_enabled	true	체크포인트 파일 생성을 활성화합니다.	true
checkpoint_store	local	local 또는 formatted_s3	"formatted_s3"
checkpoint_upload_fail_mode	warn	warn 또는 fail	"fail"
training_checkpoint_reuse_mode	off	off/verify_only/resume_if_exists	"resume_if_exists"
checkpoint_reuse_mode	off	하위 호환성 별칭	"verify_only"

[[8-8-manual-archive-parameters]]

== 8.8 수동 아카이브 매개변수

이 그룹을 사용하여 StorageGRID 아카이빙을 활성화하고, 대상 버킷 및 자격 증명을 지정하고, 아카이빙 지점을 선택할 수 있습니다. 선택한 `manual\_archive\_stage` 아카이빙 지점은 아카이빙 시점과 아카이빙에 기존 학습 아티팩트만 포함할지 또는 전체 추론 결과 세트를 포함할지를 제어합니다.

매개변수	기본값	필수	설명	예
manual_archive_enabled	false	아니요	아카이빙을 활성화합니다	true
manual_archive_stage	inferencing	아니요	보관 지점 및 아티팩트 세트: model_training 핵심 학습 아티팩트용 또는 inferencing 전체 결과 세트용	"model_training"
manual_archive_bucket	archivalbucket	활성화된 경우 예	아카이브 버킷	"archivalbucket"

매개변수	기본값	필수	설명	예
manual_archive_prefix	ai-models	아니요	아카이브 접두사	"ai-models"
manual_archive_endpoint	기본	맞춤 제작만 가능	S3 엔드포인트 아카이브	"http://10.63.150.62:10444"
manual_archive_profile	sgdlocal	권장	자격 증명 프로필	"sgdlocal"
manual_archive_profiles	None	선택 사항	프로필 맵	{"sgdlocal": {...}}
manual_archive_include_xcp_inputs	false	아니요	아카이브에 XCP 입력 포함	true
manual_archive_cleanup_raw	false	아니요	업로드 후 로컬 raw 폴더 제거	true

별칭: fabricpool_archive_* 모든 manual_archive_* 매개변수에 대해.

[[8-8-1-manual-archive-stage-decision]]

=== 8.8.1 수동 아카이브 단계 결정

The route_manual_archive_stage 분기 작업은 manual_archive_stage DAG 실행당 한 번씩 읽습니다. 선택된 값은 아카이빙 실행 시점과 StorageGRID 아카이빙 버킷에 업로드되는 아티팩트를 모두 결정합니다. 아카이브 오브젝트는 s3://<manual_archive_bucket>/<manual_archive_prefix>/<stage>/<run_stamp>/아래에 저장됩니다.

manual_archive_stage 값	아카이브 타이밍	아티팩트 아카이브됨	권장 사용법
model_training	model_training 직후에 실행되며, 미세 조정이나 추론을 기다리지 않습니다.	regression_model.bin, text_vectorizer.bin, text_classifier.bin, train_metrics.json	재현 가능한 기본 구성 모델을 유지하거나, 아카이브 용량을 최소화하거나, 이후 단계 전에 체크포인트를 보존합니다.
inferencing (기본값)	fine_tuning 및 inferencing 완료 후	모든 model_training 아티팩트 및 text_classifier_tuned.bin, fine_tune_metrics.json, tabular_predictions.csv, text_predictions.json	모델 실행에 대한 모든 비즈니스 결과 및 예측 증거를 보관합니다.

두 가지 선택 사항 중 하나를 선택하더라도, `manual\_archive\_include\_xcp\_inputs=true`를 설정하여 로컬에 구체화된 XCP 학습 입력이 있는 경우 추가합니다. `manual\_archive\_enabled=false`를 설정하여 학습 및 추론 워크플로를

변경하지 않고 아카이빙 작업을 건너뛰니다.

[[8-9-complete-sample-configuration]]

== 8.9 전체 샘플 구성

이 카탈로그는 주요 배포 및 테스트 경로에 대한 시작 구성을 제공합니다. 모든 예제는 원시 데이터가 StorageGRID에 있고 준비된 데이터가 ONTAP NAS 버킷에 기록된다고 가정합니다. 버킷 이름, 엔드포인트 주소, 파일 시스템 경로 및 프로필 이름을 대상 환경의 값으로 교체하십시오. 프로덕션 액세스 키 또는 비밀 키를 `CONF\_JSON`에 저장하지 마십시오. Airflow Connections, Variables 또는 비밀 키 백엔드를 통해 구성하십시오.

[[8-9-1-xcp-to-lustrefs-with-spark-and-delta-lake]]

=== 8.9.1 XCP에서 Spark 및 Delta Lake를 사용한 LustreFS로의 변환

이 최대 처리량 예제를 사용하여 분산 준비 및 고성능 LustreFS 모델 I/O를 구현하십시오. 이 예제는 s3_raw_prefix 아래의 모든 적합한 테이블 형식 파일을 처리하고, 준비 중에 Delta 테이블을 기록하며, 학습 체크포인트를 ONTAP NAS 준비 데이터 버킷에 저장하고, 전체 추론 결과 세트를 아카이브합니다.

```
{
  "ingestion_tool": "s3_direct",
  "s3_raw_bucket": "bucket1",
  "s3_raw_prefix": "example_ai_pipeline/raw",
  "s3_formatted_bucket": "prepnasbucket",
  "s3_formatted_prefix": "example_ai_pipeline/formatted",

  "enable_xcp": true,
  "xcp_copy_destination": "lustrefs",
  "xcp_nfs_source_ip": "10.63.150.159",
  "xcp_nfs_source_path": "/prepnasbucket/example_ai_pipeline",
  "xcp_lustrefs_source_path": "/prepnasbucket/example_ai_pipeline",
  "xcp_lustrefs_dest_path": "/mnt/lustre/client",
  "xcp_lustrefs_training_subpath": "example_ai_pipeline",
  "xcp_text_allow_local_fallback": false,

  "transformation_engine": "spark",
  "spark_required": true,
  "table_format": "delta",
  "table_format_required": true,
  "spark_driver_memory": "8g",
  "spark_executor_memory": "8g",
  "spark_timeout_secs": 0,
  "sample_count": 14400,
  "seed": 11,

  "infer_all_splits": true,
  "infer_all_text": true,

  "checkpoint_enabled": true,
  "checkpoint_store": "formatted_s3",
```

```

"training_checkpoint_reuse_mode": "off",

"manual_archive_enabled": true,
"manual_archive_stage": "inferencing",
"manual_archive_bucket": "archivalbucket",
"manual_archive_prefix": "ai-models",
"manual_archive_profile": "sgdlocal",
"manual_archive_cleanup_raw": true
}

```

[[8-9-2-no-xcp-with-python-preparation]]

=== 8.9.2 Python 준비 없이 XCP 사용

이 기본 구성을 간단한 기능 실행에 사용합니다. 준비 단계에서는 로컬 Python 엔진을 사용하고, 후속 단계에서는 파이프라인의 로컬 준비 데이터 및 아티팩트 경로를 사용합니다. XCP, Spark, 테이블 형식, 체크포인트 재사용 및 아카이빙은 비활성화되어 있습니다.

```

{
  "ingestion_tool": "s3_direct",
  "s3_raw_bucket": "bucket1",
  "s3_raw_prefix": "example_ai_pipeline/raw",
  "s3_formatted_bucket": "prepnasbucket",
  "s3_formatted_prefix": "example_ai_pipeline/formatted",
  "enable_xcp": false,
  "transformation_engine": "python",
  "table_format": "none",
  "sample_count": 200,
  "seed": 11,
  "checkpoint_enabled": false,
  "manual_archive_enabled": false
}

```

[[8-9-3-python-preparation-with-two-explicit-input-files]]

=== 8.9.3 두 개의 명시적인 입력 파일을 사용한 Python 준비

알려진 두 개의 StorageGRID CSV 오브젝트를 대상으로 파이프라인을 검증할 때 이 집중 테스트를 사용하십시오. s3_tabular_object_keys 자동 표 형식 파일 검색이 비활성화되지만, 텍스트 입력은 여전히 원시 접두사 아래의 예상 위치를 사용합니다.

```

{
  "ingestion_tool": "s3_direct",
  "s3_raw_bucket": "bucket1",
  "s3_raw_prefix": "example_ai_pipeline/raw",
  "s3_tabular_object_keys": [
    "example_ai_pipeline/raw/tabular_part_01.csv",
    "example_ai_pipeline/raw/tabular_part_02.csv"
  ]
}

```

```

],
"s3_formatted_bucket": "prepnasbucket",
"s3_formatted_prefix": "example_ai_pipeline/formatted",
"enable_xcp": false,
"transformation_engine": "python",
"table_format": "none",
"sample_count": 200,
"seed": 11,
"manual_archive_enabled": false
}

```

[[8-9-4-python-preparation-with-automatic-all-file-discovery]]

=== 8.9.4 자동 전체 파일 검색 기능을 갖춘 Python 준비

`s3+\_+\_tabular+\_+\_object+\_+\_keys` 및
`s3+\_+\_tabular+\_+\_object+\_+\_key`를 생략하면 StorageGRID 원시 데이터 접두사
아래에 있는 모든 적격 테이블 형식 오브젝트를 처리합니다. 이 예제는 Python 준비 엔진을
사용하는 모든 파일에 대한 기능 또는 확장성 테스트에 적합합니다.

```

{
"s3_raw_bucket": "bucket1",
"s3_raw_prefix": "example_ai_pipeline/raw",
"s3_formatted_bucket": "prepnasbucket",
"s3_formatted_prefix": "example_ai_pipeline/formatted",
"enable_xcp": false,
"transformation_engine": "python",
"table_format": "none",
"sample_count": 0,
"seed": 11,
"manual_archive_enabled": false
}

```

sample_count: 0 데이터 준비 과정에서 사용 가능한 모든 행을 사용한다는 의미입니다. 제한된 테스트 실행을 위해
양수 값을 설정하십시오.

[[8-9-5-xcp-to-ontap-s3-with-python-preparation]]

=== 8.9.5 Python을 사용하여 ONTAP S3로 XCP 준비

준비된 데이터를 ONTAP NAS NFS 내보내기에서 ONTAP S3 학습 버킷으로 복사해야 하는 경우 이 구성을
사용하십시오. 모델 학습, 미세 조정 및 추론 결과는 동일한 ONTAP S3 대상을 통해 생성 및 게시됩니다.

```

{
"s3_raw_bucket": "bucket1",
"s3_raw_prefix": "example_ai_pipeline/raw",

```

```

"s3_formatted_bucket": "prepnasbucket",
"s3_formatted_prefix": "example_ai_pipeline/formatted",
"enable_xcp": true,
"xcp_copy_destination": "s3",
"xcp_nfs_source_ip": "10.63.150.159",
"xcp_nfs_source_path": "/prepnasbucket/example_ai_pipeline",
"xcp_s3_endpoint": "https://ontap-s3.example.com",
"xcp_s3_bucket": "trainingbucket",
"xcp_s3_dest_prefix": "example_ai_pipeline",
"xcp_s3_training_prefix": "example_ai_pipeline",
"xcp_s3_profile": "ontaps3",
"transformation_engine": "python",
"table_format": "none",
"sample_count": 14400,
"seed": 11,
"checkpoint_enabled": true,
"checkpoint_store": "formatted_s3",
"manual_archive_enabled": true,
"manual_archive_stage": "model_training",
"manual_archive_bucket": "archivalbucket",
"manual_archive_profile": "storagegrid-archive"
}

```

[[8-9-6-xcp-to-ontap-s3-with-spark-and-iceberg]]

=== 8.9.6 Spark 및 Iceberg를 사용한 XCP에서 ONTAP S3로의 변환

Iceberg 테이블 출력 및 ONTAP S3를 활성 학습 대상으로 사용하는 분산 준비에 이 구성을 사용하십시오. Spark 또는 Iceberg를 사용할 수 없을 때 대체 실행 대신 실패해야 하는 경우 Set spark_required 및 'table_format_required'을 'true'으로 설정하십시오.

```

{
  "s3_raw_bucket": "bucket1",
  "s3_raw_prefix": "example_ai_pipeline/raw",
  "s3_formatted_bucket": "prepnasbucket",
  "s3_formatted_prefix": "example_ai_pipeline/formatted",
  "enable_xcp": true,
  "xcp_copy_destination": "s3",
  "xcp_nfs_source_ip": "10.63.150.159",
  "xcp_nfs_source_path": "/prepnasbucket/example_ai_pipeline",
  "xcp_s3_bucket": "trainingbucket",
  "xcp_s3_dest_prefix": "example_ai_pipeline",
  "xcp_s3_profile": "ontaps3",
  "transformation_engine": "spark",
  "spark_required": true,
  "table_format": "iceberg",
  "table_format_required": true,
}

```

```

"spark_driver_memory": "8g",
"spark_executor_memory": "8g",
"spark_timeout_secs": 0,
"sample_count": 0,
"seed": 11,
"manual_archive_enabled": true,
"manual_archive_stage": "inferencing",
"manual_archive_bucket": "archivalbucket",
"manual_archive_profile": "storagegrid-archive"
}

```

[[8-9-7-scenario-selection-summary]]

=== 8.9.7 시나리오 선택 요약

대본	XCP 대상	준비 엔진	표 형식	입력 범위	아카이브 단계
8.9.1	LustreFS	불꽃	Delta	발견된 모든 파일	inferencing
8.9.2	None	파이썬	None	발견된 모든 파일	비활성화됨
8.9.3	None	파이썬	None	두 개의 명시적인 파일	비활성화됨
8.9.4	None	파이썬	None	발견된 모든 파일	비활성화됨
8.9.5	ONTAP S3	파이썬	None	발견된 모든 파일	model_training
8.9.6	ONTAP S3	불꽃	Iceberg	발견된 모든 파일	inferencing

[[8-9-8-execution-walkthroughs]]

=== 8.9.8 실행 연습

Airflow 워크스페이스에서 예제를 실행합니다. 트리거 스크립트는 고유한 `manual__<UTC timestamp>` 실행 ID를 생성하고 완료될 때까지 기다린 후, 실패하거나 시간 초과된 경우 0이 아닌 종료 코드를 반환합니다. 다음 전체 구성은 운영 예제와 동일한 `CONF_JSON="$ (python3 - <<'PY' ...)"` 패턴을 사용합니다. 실행하기 전에 참조된 Airflow 관리 연결, XCP 프로파일 및 아카이브 프로ファイルを 시크릿 백엔드에서 구성하고, `CONF_JSON` 시크릿이 아닌 버킷, 엔드포인트, 경로 및 프로파일 식별자로 제한합니다.

유효한 테이블 형식 이름은 `iceberg`입니다. 지원되지 않는 값인 `icerberg`을 사용하지 마십시오.

모든 파일, **Spark**, **Iceberg**, **XCP**를 **ONTAP S3**로, 체크포인트를 **prepared-data S3**로, 추론 후 아카이브:

```

CONF_JSON="$ (python3 - <<'PY'
import json

print(json.dumps({
    "ingestion_tool": "s3_direct",
    "enable_xcp": True,
    "xcp_copy_destination": "s3",
    "xcp_text_allow_local_fallback": True,

```

```

"s3_raw_bucket": "bucket1",
"s3_raw_prefix": "example_ai_pipeline/raw",
"s3_raw_endpoint_url": "https://storagegrid.example.com",
"s3_raw_region": "us-east-1",

"s3_formatted_bucket": "prepnasbucket",
"s3_formatted_prefix": "example_ai_pipeline/formatted",
"s3_formatted_endpoint_url": "https://ontap-nas.example.com",
"s3_formatted_region": "us-east-1",

"xcp_nfs_source_ip": "10.63.150.159",
"xcp_nfs_source_path": "/prepnasbucket/example_ai_pipeline",
"xcp_s3_endpoint": "https://ontap-s3.example.com",
"xcp_s3_bucket": "trainingbucket",
"xcp_s3_dest_prefix": "example_ai_pipeline",
"xcp_s3_training_prefix": "example_ai_pipeline",
"xcp_s3_profile": "ontaps3",

"transformation_engine": "spark",
"spark_required": True,
"table_format": "iceberg",
"table_format_required": True,
"sample_count": 0,
"seed": 11,
"spark_driver_memory": "8g",
"spark_executor_memory": "8g",
"spark_timeout_secs": 0,
"infer_all_splits": True,
"infer_all_text": True,

"checkpoint_enabled": True,
"checkpoint_store": "formatted_s3",
"manual_archive_enabled": True,
"manual_archive_stage": "inferencing",
"manual_archive_bucket": "archivalbucket",
"manual_archive_prefix": "ai-models",
"manual_archive_profile": "sgdlocal",
"manual_archive_endpoint": "https://storagegrid.example.com",
"manual_archive_cleanup_raw": True,
"manual_archive_include_xcp_inputs": True,
}, separators=(",", ":"))
PY
)"
CONF_JSON="$CONF_JSON" AIRFLOW_BIN=./.venv/bin/airflow INGESTION_TOOL
=s3_direct \

```

```
./trigger_and_wait_ai_pipeline_sklearn.sh
```

모든 파일, **Spark**, **Iceberg**, **XCP to LustreFS**, **checkpoint to prepared-data S3**, 추론 후 아카이브:

```
CONF_JSON="$(python3 - <<'PY'
import json

print(json.dumps({
    "ingestion_tool": "s3_direct",
    "enable_xcp": True,
    "xcp_copy_destination": "lustrefs",
    "ssh_user": "root",
    "ssh_host": "10.63.150.178",
    "xcp_text_allow_local_fallback": True,

    "s3_raw_bucket": "bucket1",
    "s3_raw_prefix": "example_ai_pipeline/raw",
    "s3_raw_endpoint_url": "https://storagegrid.example.com",
    "s3_raw_region": "us-east-1",

    "s3_formatted_bucket": "prepnasbucket",
    "s3_formatted_prefix": "example_ai_pipeline/formatted",
    "s3_formatted_endpoint_url": "https://ontap-nas.example.com",
    "s3_formatted_region": "us-east-1",

    "xcp_nfs_source_ip": "10.63.150.159",
    "xcp_nfs_source_path": "/prepnasbucket/example_ai_pipeline",
    "xcp_lustrefs_source_path": "/prepnasbucket/example_ai_pipeline",
    "xcp_lustrefs_dest_path": "/mnt/lustre/client",
    "xcp_lustrefs_training_subpath": "example_ai_pipeline",

    "transformation_engine": "spark",
    "spark_required": True,
    "table_format": "iceberg",
    "table_format_required": True,
    "sample_count": 0,
    "seed": 11,
    "spark_driver_memory": "8g",
    "spark_executor_memory": "8g",
    "spark_timeout_secs": 0,
    "infer_all_splits": True,
    "infer_all_text": True,

    "checkpoint_enabled": True,
    "checkpoint_store": "formatted_s3",
```

```

    "manual_archive_enabled": True,
    "manual_archive_stage": "inferencing",
    "manual_archive_bucket": "archivalbucket",
    "manual_archive_prefix": "ai-models",
    "manual_archive_profile": "sgdlocal",
    "manual_archive_endpoint": "https://storagegrid.example.com",
    "manual_archive_cleanup_raw": True,
    "manual_archive_include_xcp_inputs": True,
}, separators=(",", ":"))
PY
)"
AIRFLOW_BIN=./.venv/bin/airflow INGESTION_TOOL=s3_direct \
./trigger_and_wait_ai_pipeline_sklearn.sh

```

ONTAP S3 모드의 경우, 비밀이 아닌 xcp_s3_profile 이름만 CONF_JSON`에 전달하고 실행 전에 XCP 호스트 또는 Airflow에서 관리하는 비밀 저장소를 통해 해당 자격 증명을 정의하십시오. LustreFS 모드의 경우, 학습 경로가 ONTAP S3 자격 증명을 확인하지 않으므로 `xcp_s3_profiles` 맵을 의도적으로 생략합니다. 두 예제 모두에서 StorageGRID 원시 데이터, ONTAP NAS 준비 데이터 및 아카이브 자격 증명을 인라인 JSON이 아닌 Airflow에서 관리하는 비밀 저장소를 통해 정의하십시오. `s3_tabular_object_keys`를 생략하고 `sample_count`를 `0`으로 설정하면 모든 적합한 테이블 형식 파일의 자동 검색 및 사용 가능한 모든 행 사용이 요청됩니다.

Python, XCP 미사용, 두 파일 테스트:

```

CONF_JSON='{ "s3_raw_bucket": "bucket1", "s3_raw_prefix": "example_ai_pipeline/
raw", "s3_tabular_object_keys": ["example_ai_pipeline/raw/tabular_part_01.c
sv", "example_ai_pipeline/raw/tabular_part_02.csv"], "s3_formatted_bucket": "
prepnasbucket", "s3_formatted_prefix": "example_ai_pipeline/formatted", "enab
le_xcp": false, "transformation_engine": "python", "table_format": "none", "samp
le_count": 200, "manual_archive_enabled": false}' \
./trigger_and_wait_ai_pipeline_sklearn.sh

```

Spark, Delta Lake, XCP에서 LustreFS로:

```

CONF_JSON='{ "enable_xcp": true, "xcp_copy_destination": "lustrefs", "xcp_nfs_s
ource_ip": "10.63.150.159", "xcp_nfs_source_path": "/prepnasbucket/example_ai
_pipeline", "xcp_lustrefs_source_path": "/prepnasbucket/example_ai_pipeline"
, "xcp_lustrefs_dest_path": "/mnt/lustre/client", "xcp_lustrefs_training_subp
ath": "example_ai_pipeline", "transformation_engine": "spark", "spark_required
": true, "table_format": "delta", "table_format_required": true, "sample_count":
14400, "manual_archive_enabled": true, "manual_archive_stage": "inferencing"}' \
./trigger_and_wait_ai_pipeline_sklearn.sh

```

Spark, Iceberg, XCP에서 ONTAP S3로:

```
CONF_JSON='{ "enable_xcp":true,"xcp_copy_destination":"s3","xcp_nfs_source_ip":"10.63.150.159","xcp_nfs_source_path":"/prepnasbucket/example_ai_pipeline","xcp_s3_bucket":"trainingbucket","xcp_s3_dest_prefix":"example_ai_pipeline","xcp_s3_profile":"ontaps3","transformation_engine":"spark","spark_required":true,"table_format":"iceberg","table_format_required":true,"sample_count":0,"manual_archive_enabled":true,"manual_archive_stage":"inferencing"}' \
./trigger_and_wait_ai_pipeline_sklearn.sh
```

XCP를 실행하기 전에 구성된 ONTAP NAS NFS 경로와 대상에 XCP 호스트에서 접근 가능한지 확인하십시오. 실행이 성공적으로 완료되면 작업 로그에서 [data_prep] effective_config, [xcp_copy] effective_config, [model_training] effective_config, 및 [manual_archive]를 검토하여 선택한 엔진, 입력 범위, 대상 및 아카이브 단계를 확인하십시오.

9. 운영 지침

Karthikeyan Nagalingam, NetApp

이 가이드는 일상적인 실행, 모니터링, 사고 분류, 교육 복구, 아카이브 검증, 용량 조정 및 데이터 보호에 대해 다룹니다. 8절의 구성 시나리오와 함께 사용하면 로컬, ONTAP S3 및 LustreFS 교육 모드 전반에 걸쳐 동일한 파이프라인을 일관되게 운영할 수 있습니다.

[[9-1-execution-walkthrough]]

== 9.1 실행 과정 안내

`trigger+\_+and+\_+wait+\_+ai+\_+pipeline+\_+sklearn.sh`를 통해 실행을 트리거하고 `CONF+\_+\_JSON` 페이로드를 사용합니다. 스크립트는 고유한 이름의 수동 실행을 한번 제출하고, 종료 상태에 도달할 때까지 폴링하며, 실패한 작업에 대한 로컬 로그 테일을 가능한 경우 출력합니다. 각 기본 작업은 시작 시 `effective+\_+\_config` 레코드를 생성하므로 운영자는 제출된 페이로드에만 의존하지 않고 평가된 구성을 확인할 수 있습니다.

[[9-2-monitoring-and-observability]]

== 9.2 모니터링 및 관찰 가능성

Airflow UI 또는 CLI에서 DAG 상태를 모니터링하고 실행 ID와 run_stamp`의 상관 관계를 확인합니다. 가드 로그는 grep으로 쉽게 검색할 수 있습니다: `[data_prep] effective_config, [model_training] effective_config, [xcp_copy] effective_config, 그리고 [checkpoint] resume_summary. 이러한 항목을 검토하여 각 실행에 대한 변환 엔진, 선택된 XCP 대상, 학습 소스 및 체크포인트 재사용 결정을 검증합니다.

[[9-3-troubleshooting-quick-reference]]

== 9.3 문제 해결 빠른 참조

이 표를 사용하여 일반적인 작업 실패와 첫 번째 해결 조치를 연결하십시오. 실행을 다시 시도하기 전에 소스 연결 및 마운트 문제를 해결하십시오. 잘못된 구성이나 사용 불가능한 엔드포인트를 수정하지 않고 다시 시도하면 동일한 오류가

발생합니다.

징후	가능성 있는 원인	해결
Could not resolve XCP S3 credentials	누락/잘못된 xcp_s3_profile 항목(xcp_s3_profiles)	프로필 이름 및 자격 증명 맵을 확인합니다
no readable Lustre source directory	LustreFS가 마운트되지 않았거나 잘못되었습니다. xcp_lustreufs_dest_path	파일 시스템 유형을 확인하세요 findmnt -T /mnt/lustre/client -o FSTYPE,TARGET -n; 올바른 경로
Missing required text dataset files	원시 S3 접두사에 없는 텍스트 JSON	업로드 text_base.json, text_finetune.json, text_infer.json
Preflight 종료 코드 1	SSH에 연결할 수 없습니다	연결 ssh_default 및 호스트 /사용자 확인
Preflight 종료 코드 2	NFS 경로가 내보내지지 않았거나 표시되지 않습니다	내보내기를 확인하세요 xcp show <ip>

[[9-4-checkpoint-operations]]

== 9.4 체크포인트 작업

체크포인트 기능은 model_training`에만 적용됩니다. 유효한 학습 완료 결과를 재사용하고 중복 모델 학습을 건너뛰려면 `training_checkpoint_reuse_mode=resume_if_exists`로 설정하거나, 건너뛰지 않고 체크포인트 적격성을 검증하려면 `verify_only`를 사용하세요. 초기 테스트 중이나 학습 입력, 구성 또는 예상 결과물이 변경된 경우에는 재사용 기능을 비활성화 상태(`off)로 유지하세요.

[[9-5-manual-archive-operations]]

== 9.5 수동 아카이브 작업

```
`manual++_++archive++_++enabled=true`을 설정하고  
`manual++_++archive++_++stage`을 신중하게 선택하십시오. 거버넌스에서 학습이 완료되는  
즉시 핵심 모델 기준선이 필요한 경우 `model++_++training`을 사용하고, 아카이브에 최종  
예측 결과를 포함해야 하는 경우 `inferencing`을 사용하십시오.  
`model++_++training`이 선택된 상태로 실행하면 아카이브 분기 이후에도 미세 조정 및 추론이  
계속 진행되지만, 이후 출력 결과는 아카이브되지 않습니다. 실행 스탬프가 찍힌 StorageGRID  
접두사로 업로드된 파일을 확인하고, 선택한 단계, 발견된 파일, 업로드 횟수에 대한  
`++[++manual++_++archive++]++` 로그를 검토하십시오.
```

[[9-6-scaling-guidance]]

== 9.6 확장 지침

```
`sample++_++count`, `spark++_++driver++_++memory`,  
`spark++_++executor++_++memory`, 및 `spark++_++timeout++_++secs`를 입력 볼륨  
및 서비스 수준 목표에 맞게 조정하십시오. 먼저 제한된 샘플로 데이터 형태와 라우팅을 검증한  
다음, StorageGRID, ONTAP NAS, XCP 및 선택한 학습 계층의 용량과 처리량이 충분해지면
```

전체 행 실행에 ``sample++_++count=0``를 사용하십시오.

[[9-7-backup-and-recovery]]

== 9.7 백업 및 복구

ONTAP 스냅샷 및 백업을 사용하여 ONTAP NAS 준비 데이터 버킷을 보호하십시오. 원시 버킷과 아카이브 버킷에 StorageGRID 보호 및 보존 정책을 적용하십시오. 아카이브 버킷은 파이프라인 재실행과 관계없이 기록을 보존합니다. 이러한 스토리지 보호 기능을 교육 체크포인트 기록과 결합하여 작업, 호스트 또는 사이트 수준의 중단 후에도 올바른 실행 범위 데이터와 기준선 아티팩트를 복구할 수 있습니다.

10. 보안 고려 사항, 검증 및 테스트

Karthikeyan Nagalingam, NetApp

이 통합 섹션에서는 AI 파이프라인을 안전하게 운영하는 데 필요한 제어 기능과 대표적인 환경에서 설계가 예상대로 작동하는지 입증하는 데 사용되는 검증 활동을 다룹니다. 파이프라인에 사용된 것과 동일한 스토리지 계층 분리, 최소 권한 접근 모델, 체크포인트 로직 및 실행 범위 기반 계보 관리가 검증 계획 및 운영 가이드라인에도 적용됩니다.

[[10-1-security-considerations]]

== 10.1 보안 고려 사항

보안 제어는 감사에 필요한 데이터 계보를 유지하면서 각 스토리지 계층과 파이프라인 기능을 격리해야 합니다. StorageGRID 원시 데이터, ONTAP NAS 준비 데이터, 선택한 XCP 대상 및 StorageGRID 아카이빙에 대해 각각 별도의 최소 권한 자격 증명을 사용하십시오. 네트워크 트래픽을 암호화하고 모든 비밀 정보는 DAG 정의, 예제 및 셀 기록 외부에 저장하십시오.

영역	권장 사항
자격 증명 관리	Airflow 연결/변수 또는 시크릿 관리자를 사용하고, 본문 내 인라인 일반 텍스트를 사용하지 마십시오. <code>dag_run.conf</code>
데이터 전송 중	TLS가 활성화된 S3 엔드포인트를 사용하고, 지원되는 경우 암호화된 NFS/Lustre 전송을 사용합니다.
액세스 제어	역할별 S3 자격 증명 범위(StorageGRID 원시 데이터, ONTAP NAS 준비 데이터, XCP 대상, StorageGRID 아카이빙)를 최소 필수 권한으로 제한합니다.
설정 파일의 비밀 정보	테스트 중에 노출된 자격 증명을 교체하고, 트리거 스크립트/셀 기록은 민감한 정보로 취급하십시오.
감사 추적	규정 준수 증거를 위해 데이터 준비 매니페스트, 체크포인트 레코드 및 아카이브 매니페스트를 활용하십시오.

[[10-2-validation-and-testing]]

== 10.2 검증 및 테스트

검증을 통해 파이프라인의 구성 가능한 동작이 라우팅, 오류 처리, 데이터 이동성 및 다중 파일 검색 경로 전반에 걸쳐 안정적으로 작동함을 확인했습니다. 각 검증 테스트 영역은 실제 Airflow DAG 트리거를 사용하는 대표적인 환경에서 실행되었으며, 로그 출력과 스토리지 매니페스트의 정확성을 검증했습니다.

[[10-2-1-functional-routing-validation]]

=== 10.2.1 기능적 라우팅 유효성 검사

이 테스트 영역은 `enable\_xcp=true`일 때 엔드투엔드 데이터 및 아티팩트 라우팅을 검증합니다.

- 목표: 데이터 준비, XCP 이동성, 모델 학습, 미세 조정 및 추론 과정에서 선택된 XCP 대상 (s3 또는 lustrefs)가 일관되게 사용되는지 확인합니다.
- 테스트 절차: `xcp_copy_destination="s3"` 및 `xcp_copy_destination="lustrefs"`로 DAG 실행을 트리거했습니다. XCom 출력, 유효 구성 로그 및 저장 위치를 검사했습니다.
- 검증 결과:
 - `data_prep`에서 형식화된 표 형식 분할 및 텍스트 파일이`
``<xcp_prefix>/formatted/<run_stamp>/data/` 아래의 ONTAP NAS 준비 데이터 계층에`
`준비되었습니다.`
 - ``xcp_copy`에서, 준비된 데이터 세트는 후속 모델 단계를 위해 선택된 XCP 대상으로 전송되었습니다.`
 - `model_training`에서 모델 입력은 XCP 대상에서 로드되었고 기존 아티팩트`
`(`regression_model.bin, text_vectorizer.bin, text_classifier.bin,`
`train_metrics.json)`는 `<xcp_prefix>/formatted/<run_stamp>/artifacts/`에 다시 게시되었습니다.`
 - `fine_tuning`에서 기본 벡터화기/분류기 아티팩트는 XCP 대상에서 구체화되고, 점진적으로`
`조정되고, `text_classifier_tuned.bin` 및 fine_tune_metrics.json(으)로 다시
게시되었습니다.`
 - 본 연구에서는 `inferencing`, 조정된 텍스트 분류기와 기존 회귀 모델 및 벡터화기를 XCP 대상에서
구체화하여 예측값을 생성했습니다.

[[10-2-2-local-fallback-and-non-xcp-execution-validation]]

=== 10.2.2 로컬 폴백 및 비XCP 실행 유효성 검사

이 테스트 영역은 XCP 데이터 이동성이 비활성화된 경우 파이프라인 동작을 검증합니다(`enable_xcp=false`).

- 목표: SSH 연결, 원격 XCP 호스트 또는 XCP 자격 증명 프로필 없이 직접 로컬 준비 데이터 경로를 사용하여 파이프라인이 원활하게 작동하는지 확인합니다.
- 테스트 절차: `enable_xcp=false` 및 기본 S3/로컬 경로를 사용하여 파이프라인 실행을 트리거했습니다.
- 검증 결과:
 - XCP 사전 검사 및 복사 작업이 안전하게 건너뛰었습니다.
 - `model_training, fine_tuning,` 및 `inferencing`은(는) 로컬에 준비된 데이터 디렉터리와 게시된
아티팩트에서 직접 읽습니다.`
 - XCP를 비활성화했을 때 XCP S3 자격 증명 확인 오류나 SSH 사전 요청 오류가 발생하지 않음을
확인했습니다.

[[10-2-3-performance-and-tier-comparison-ontap-s3-vs-lustrefs]]

=== 10.2.3 성능 및 계층 비교 (ONTAP S3 vs. LustreFS)

이 테스트 영역은 ONTAP S3 오브젝트 스토리지와 LustreFS 병렬 파일 시스템 계층 간의 I/O 지연 시간, 검색 오버헤드 및 학습 처리량을 평가합니다.

- 목표: LustreFS를 사용하는 E-Series와 ONTAP S3 활성 학습 계층을 사용하는 ONTAP AFF/AFX의 성능 특성을 정량화합니다.
- 테스트 절차: 동일한 데이터셋 크기(14,400+ 행, 다중 파일 텍스트/테이블 형식 입력)에서 파일 검색 시간, 학습

데이터셋 로드 지연 시간 및 아티팩트 게시/구체화 기간을 측정했습니다.

- 검증 결과:

- **LustreFS** 티어: 모델 학습 중 파일 검색 오버헤드가 낮고 임의 액세스 읽기 지연 시간이 더 빠른 것으로 입증되어 파일 수가 많고 GPU 집약적인 병렬 학습 워크로드에 이상적입니다.
- **ONTAP S3 Tier**: 간편한 멀티 프로토콜 관리로 높은 처리량을 제공하여 클라이언트 측 파일 시스템 마운트 없이도 비용 효율적이고 탄력적인 액세스가 가능한 액티브 스토리지에 최적입니다.

[[10-2-4-failure-recovery-and-training-checkpoint-validation]]

=== 10.2.4 장애 복구 및 훈련 체크포인트 검증

이 테스트 영역은 `model\_training`로 범위가 지정된 체크포인트 재개 시스템의 유효성을 검증합니다.

- 목표: 파이프라인 복구가 이전에 성공적으로 완료된 학습 실행을 정확하게 감지하고 아티팩트 무결성을 유지하면서 중복 계산을 건너뛰는지 확인합니다.
- 테스트 절차:
 - a. `checkpoint\_enabled=true`와 `checkpoint\_store="formatted\_s3"`를 사용하여 파이프라인을 실행했습니다.
 - b. 시뮬레이션된 작업 실패 또는 재실행(`training\_checkpoint\_reuse\_mode="resume\_if\_exists"` 포함).
 - c. 테스트 `verify_only` 및 `off` 재사용 모드.
- 검증 결과:
 - `resume\_if\_exists`가 설정되고 유효한 체크포인트 JSON + 모든 핵심 아티팩트가 S3/LustreFS에 존재하면, `model\_training`는 결정 `RESUMED\_FROM\_CHECKPOINT`과 함께 즉시 종료되고 `[checkpoint] resume\_summary: decision=RESUMED\_FROM\_CHECKPOINT`를 기록합니다.
 - 다운스트림 `fine_tuning` 및 `inferencing` 검증된 체크포인트 아티팩트를 성공적으로 구체화했습니다.
 - `verify\_only`이(가) 설정되면, 가드는 학습을 건너뛰지 않고 아티팩트의 존재 여부를 검증했습니다.

[[10-2-5-scalability-and-multi-file-dataset-validation]]

=== 10.2.5 확장성 및 다중 파일 데이터셋 유효성 검사

이 테스트 영역은 대규모의 다중 파일로 구성된 표 형식 및 텍스트 데이터 세트 전반에 걸쳐 파이프라인의 확장성을 검증합니다.

- 목표: 자동 데이터셋 검색, Spark 변환 엔진 실행 및 Lakehouse 테이블 형식 지원 (delta 및 iceberg)을 확인합니다.
- 테스트 절차:
 - a. StorageGRID 원시 데이터 접두사 아래에 있는 수십 개의 CSV/Parquet 파일에 대한 데이터 수집 및 준비 작업을 수행했습니다(`s3_raw_prefix`).
 - b. 명시적 파일 선택 방식(`s3_tabular_object_keys`) 대 자동 전체 파일 검색 방식(`sample_count=0`)을 테스트했습니다.
 - c. PySpark을 사용하여 `table_format="delta"` 및 `table_format="iceberg"`로 준비를 실행했습니다.
- 검증 결과:
 - Spark 분산 준비 기능을 사용하여 대규모 다중 파트 테이블형 데이터 세트를 성공적으로 수집, 필터링, 분할 및 포맷했습니다.
 - 자동 검색 기능은 비테이블 형식의 아티팩트를 무시하면서 모든 유효한 테이블 형식 CSV/Parquet 오브젝트를

정확하게 식별했습니다.

- Delta Lake와 Iceberg 테이블 형식 모두 올바르게 생성되어 `tables/`에 등록되었으며, 다운스트림 모델 훈련에서 형식화된 분할을 검색하고 읽었습니다.

[[10-2-6-example-customer-evaluation-workflow-sizing-assumptions-and-validation-metrics]]

=== 10.2.6 고객 평가 워크플로 예시, 규모 산정 가정 및 검증 지표

이 워크플로는 고객이 설계가 AI 파이프라인 성숙도, 데이터 주권 요구 사항 및 성능 목표에 적합한지 평가하는 데 도움이 됩니다. 대표 데이터 세트와 파이프라인 실행 1회로 시작하여 각 승인 기준이 충족된 후에만 데이터 볼륨, 파일 수, 모델 복잡성 및 동시 실행 횟수를 늘립니다. ONTAP S3와 LustreFS 간의 비교에서 동일한 입력 데이터와 구성을 사용할 수 있도록 `run\_stamp`별로 결과를 기록합니다.

평가 단계	워크플로우 예시	규모 산정 가정/결정	검증 지표 및 수용 증거
1. 기준선 설정	제한된 샘플에서 Python 준비 경로를 <code>enable_xcp=false</code> 실행합니다.	섹션 7.1.1의 참조 검증 환경을 초기 기능 기준선으로 사용하십시오.	DAG 완료 성공 여부; 입력 행 및 출력 분할 횟수; 모델 지표; 작업 시간; CPU, 메모리 및 로컬 디스크 사용률.
2. 데이터 주권 검증	원시 데이터, 준비된 데이터, 활성 학습 데이터 및 아카이브는 승인된 스토리지 엔드포인트 및 지역에 보관하십시오. 계층별 자격 증명 및 보존 정책을 검토하십시오.	데이터를 이동하기 전에 허용된 위치, 액세스 역할, 암호화 요구 사항 및 보존 정책을 정의하십시오.	엔드포인트, 버킷, 접두사 및 리전이 유효한 구성 및 매니페스트에 기록됩니다. 최소 권한 액세스 검증이 이루어지며, 아카이브 및 삭제 정책에 대한 증거가 제공됩니다.
3. 데이터 이동성 검증	XCP를 활성화하고 준비된 실행 결과를 ONTAP S3에 복사한 다음, LustreFS에도 동일한 작업을 반복합니다.	준비된 실행, 아티팩트, 작업 공간 및 동시 실행을 위해 10 GbE 연결과 충분한 대상 용량을 확보하십시오.	XCP 완료 상태; 복사된 바이트 및 파일 수; 복사 시간 및 처리량; 소스/대상 파일 수 및 체크섬 또는 매니페스트 비교; 사전 검사 성공 여부.
4. 훈련 단계 적합성 검증	선택한 각 목적지에 대해 동일한 데이터 세트와 모델 설정을 사용하여 학습, 미세 조정 및 추론을 수행합니다.	객체 기반 워크플로우에는 ONTAP S3를 사용하고, 병렬 학습 I/O 또는 높은 파일 수가 정당화되는 경우에는 LustreFS와 함께 E-Series를 사용하십시오.	데이터셋 검색 및 로드 시간; 학습, 미세 조정 및 추론 시간; 아티팩트 게시/실행 시간; 해당되는 경우 CPU/GPU 사용률; 모델 품질 일관성.
5. 규모 및 복구 검증	모든 행으로 `sample_count` 확장하고, 파일 수와 동시 실행 횟수를 늘린 다음, 체크포인트 재사용 및 아카이빙을 테스트하십시오.	유지되는 실행 범위 데이터 세트 및 아티팩트의 크기 용량과 향후 성장을 위한 여유 공간, Spark 및 동시 Airflow 작업에 필요한 CPU 및 메모리 크기를 산정합니다.	전체 소요 시간; 성공률; 대기 시간; 체크포인트 재개 결정 및 소요 시간; 아카이브 완전성; 실행당 저장 공간 증가량; 복구 시간 목표(RTO) 달성률.

평가를 시작하기 전에 고객은 엔드투엔드 실행 시간, XCP 처리량, 학습 데이터 로드 시간, 최대 동시 실행 수, 복구 시간 및 스토리지 보존 기간에 대한 정량적 목표를 설정해야 합니다. 측정된 기준선과 각 확장 테스트를 이러한 목표와 비교하여 아키텍처가 의도된 워크로드 및 운영 요구 사항을 충족하는지 확인해야 합니다.

11. NetApp 차별화 및 경쟁력, 부록

Karthikeyan Nagalingam, NetApp

이 통합 섹션에서는 아키텍처를 일반적인 대안과 비교하여 설명하고 운영자와 검토자가 공유하는 참조 자료를 제공합니다. 차별점은 단일 스토리지 엔드포인트가 아니라, StorageGRID, ONTAP NAS, ONTAP S3, LustreFS, XCP 및 Airflow의 정책 제어형 조합으로, 각 AI 라이프사이클 단계에서 워크플로를 재구축하지 않고도 적절한 계층을 사용할 수 있게 합니다.

[[11-1-netapp-differentiation-and-competitive-positioning]]
== 11.1 NetApp 차별화 및 경쟁력 강화 전략

이 비교는 해당 아키텍처를 일반적인 대안들과 비교하여 보여줍니다. 차별점은 단일 스토리지 엔드포인트가 아니라, StorageGRID, ONTAP NAS, ONTAP S3, LustreFS, XCP 및 Airflow의 정책 제어형 조합으로, 각 AI 라이프사이클 단계에서 워크플로를 재구축하지 않고도 적절한 계층을 사용할 수 있게 해줍니다.

대안적 접근법	제한 사항	NetApp 솔루션의 장점
클라우드 네이티브 전용 파이프라인	하나의 스토리지 패러다임에 고정됨	구성을 통해 S3/LustreFS 상호 교환 가능
DIY 스크립트 기반 파이프라인	재시도, 체크포인트 또는 계보 없음	Airflow 오케스트레이션 + 체크포인트/재개 기능을 기본적으로 지원합니다.
단일 계층 스토리지 아키텍처	모든 워크로드에 적용 가능한 단일 성능/비용 프로필	실행별 계층 선택(via xcp_copy_destination
모든 업데이트에 대한 전체 재교육	높은 반복 컴퓨팅 비용	~을 통한 점진적 미세 조정 partial_fit

[[11-2-appendices]]
== 11.2 부록

부록은 이 NVA를 운영, 검토 또는 수정하는 사용자를 위해 공통 용어 및 문서 관리 맥락을 제공합니다. 문서화된 동작, 매개변수, 저장 레이아웃 또는 유효성 검사 증거를 변경할 때는 개정 이력을 최신 상태로 유지하십시오.

[[11-2-1-glossary]]
=== 11.2.1 용어집

이러한 정의는 아키텍처의 저장, 오케스트레이션 및 계보 용어에 대한 일관된 의미를 확립합니다.

용어	정의
엑스피	NetApp의 고속 병렬 데이터 복사/마이그레이션 툴
ONTAP S3	NetApp의 ONTAP 기반 S3 호환 오브젝트 스토리지 서비스
LustreFS	고성능 병렬 분산 파일 시스템
DAG	방향성 비순환 그래프 — Airflow의 워크플로 정의
XCom	Airflow에서 태스크 간 데이터 전달을 위한 메커니즘
run_stamp	계보에 사용되는 실행별 고유 타임스탬프 식별자

용어	정의
명백한	스테이지의 입력, 출력 및 구성을 설명하는 JSON 레코드

[[11-2-2-revision-history]]

=== 11.2.2 개정 내역

문서의 검토 및 승인 내역을 보존하기 위해 게시된 모든 변경 사항을 기록하십시오.

버전	날짜	작성자	변경
1.0	2026-09-25	솔루션 아키텍처 팀	전체 매개변수 표와 도표를 포함한 초기 출판용 NVA

[[11-2-3-contact-us-and-software-requests]]

=== 11.2.3 문의 및 소프트웨어 요청

이 아키텍처를 배포하고 검증하는 데 필요한 소프트웨어 패키지, 자동화 DAG, 배포 스크립트 또는 검증 아티팩트를 요청하거나 피드백 및 기술 문의 사항이 있는 경우 NetApp AI 및 데이터 모빌리티 팀에 문의하십시오.

- 이메일: ng-data-mobility-in-ai-pipeline@netapp.com
- **Purpose:** 소프트웨어 패키지 요청, 배포 지침, 아키텍처 검증 지원 및 기술 피드백 제공.

저작권 정보

Copyright © 2026 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.