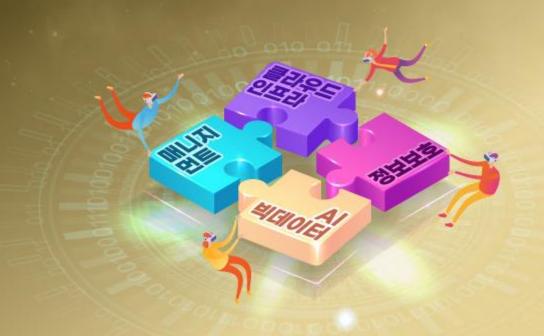


차세대데이터 플랫폼 CDP 소개

이래훈 이사 굿모닝아이텍㈜



AGENDA

- CDP(Cloudera Data Platform)의 이해
- CDP HYBRID CLOUD



CDP(Cloudera Data Platform)의 이해



CLOUDERA | THE ENTERPRISE DATA CLOUD COMPANY

모든 Cloud 및 Data center에서 End-To-End 데이터 분석 플랫폼 제공





Security | Governance | Lineage | Management | Automation



Architecture & Technology의 진화

CLOUD EXPERIENCE



- Easy to use
- Self-service
- On-demand
- Elastic
- Consumption 모델

COMPUTE & STORAGE



- Public & Private Cloud
- 성능 및 운영 안정성을 위해
 Compute와 Storage
 Layer를 분리하여 활용

KUBERNETES & CONTAINERS



• 유연성과 민첩성을 위한 표준 운영환경으로 Cloud Platform 채택

STREAMING & ML/AI



 데이터 기반의 기업으로 변화하기 위해
 Multi-function Analytics 활용



Enterprise 환경에서 요구되는 기능





- 멀티 클라우드
- 하이브리드 클라우드
- 프라이빗 클라우드
- 데이터센터
- 에지 디바이스



Data Lifecycle

- 스트리밍
- 데이터 엔지니어링
- 데이터 웨어하우스
- 머신 러닝 & 인공 지능



Secure & Governed

- 데이터 품질 & 메타데이터
- 상세 수준의 보안 정책
- 데이터 계보관리
- 데이터 & 워크로드 마이그레이션



Open, DataOps

- 100% 오픈 소스
- 오픈 데이터 형식
- 오픈 스토리지 및 컴퓨팅
- 오픈 APIs



Cloudera Product Portfolio

Edge, Public/Private/Hybrid Cloud, On Premise의 모든 환경에서 통합적인 데이터 플랫폼 제공

Cloudera Data Platform (CDP)





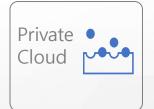












Edge에서 AI 영역까지 통합된 데이터 관리 및 분석 플랫폼



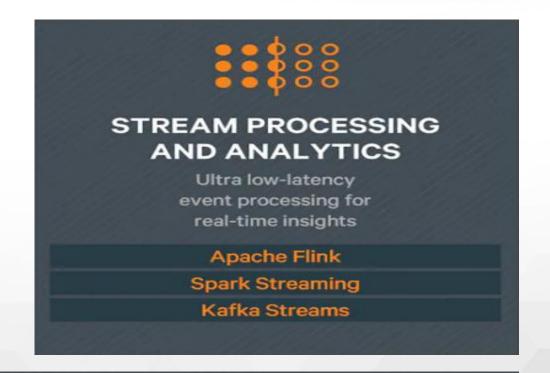






Cloudera Data Flow Component





STREAMS MESSAGING

High-speed streams messaging, monitoring and replication



Apache Kafka

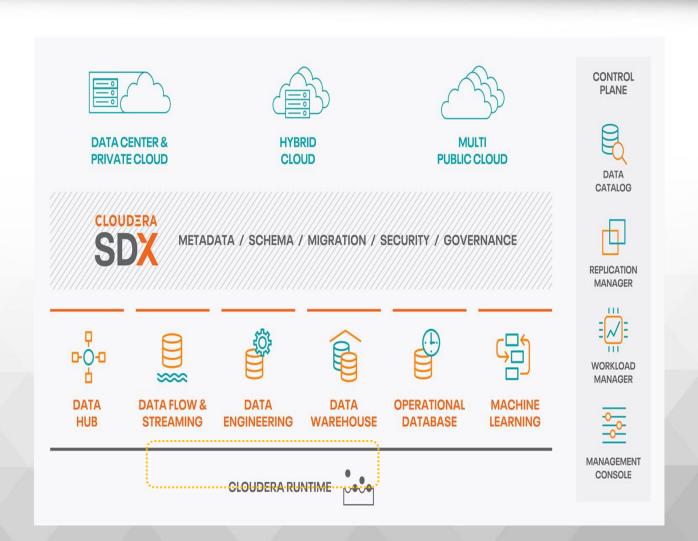
Streams Messaging Manager

Streams Replication Manager



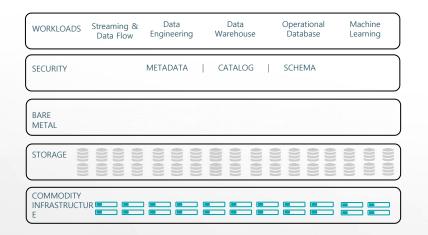
Cloudera Data Platform

- 최적의 Cloud 비용을 위한 자동 확장, 자동 중지 및 자동 실행 기능 제공 (auto scale, suspend and resume)
- 데이터 분석 및 AI/ML 분석 워크로드 최적을 위한 환경 제공
- 영구 및 임시 클러스터에 대한 종합적인 데이터 거버넌스 기능 제공 : Data lineage
- 단일 관리 도구에서 Hybrid 및 Multi-cloud 환경 내의 분석 플랫폼 통합 관리
- 뛰어난 확장성 제공 : Petabytes 데이터 및 수천 명 이상의 동시 사용자 수용 가능

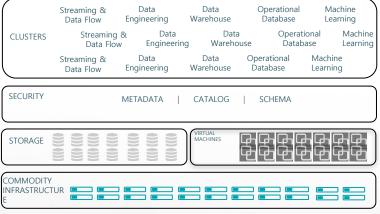


3세대 아키텍처로 진화

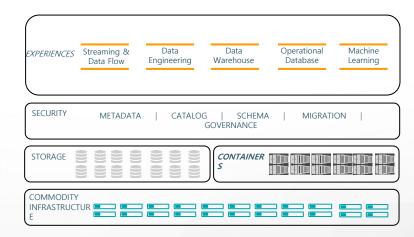
On Premise



Public Cloud



Multi-Public/Private/Hybrid Cloud



Co-located Storage/Compute Large, Shared Clusters (CDH / HDP) Disaggregated Storage/Compute Multiple Clusters (EMR / HDInsight) Disaggregated Storage/Compute Multi-Tenant, Containerized "SaaS Experiences" (CDP)

Gen#1

Gen#2

Gen#3

Evolution

HYBRID DATA CLOUD



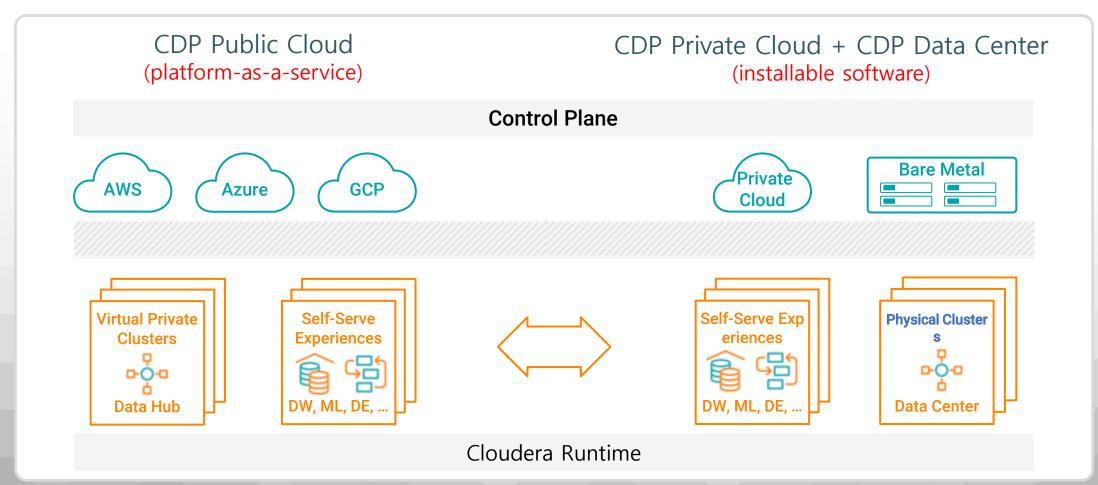
"The reality: Hybrid cloud is The de facto model."

Deloitte.



CDP를통해엔터프라이즈데이터클라우드지원

CDP (Cloudera Data Platform)



CDP-Data Center가 지원하는 3가지 Deployment Models

Traditional/Split/Hybrid

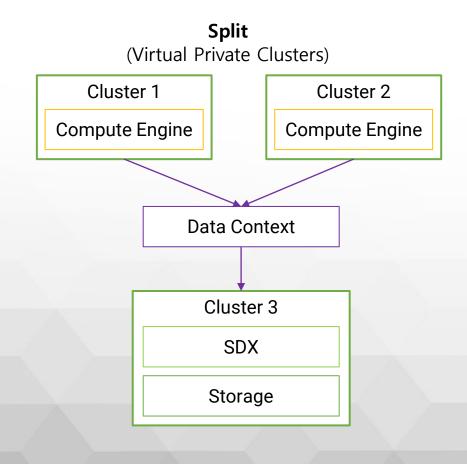
Traditional (Data Locality)

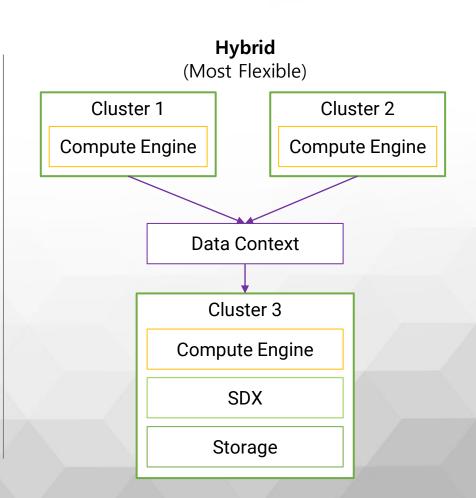
Cluster

Compute Engine

SDX

Storage





CDP Data Center 개요





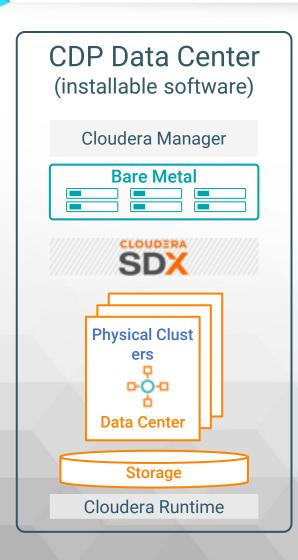
IN-PLACE UPGRADE



CDH & HDP의 최적의 조합 합통합된 최종 버전 (Cloudera Runtime) 다양한 운영 환경에서 입증된 아키텍처 (servers)

CDH / HDP에서 업그레이드 지원 (in-place upgrade) Private Cloud를 활용하기 위한 기본 조건 (Cloudera Runtime)

CDP Data Center개요



CDP Data Center 신 기능

- HDP + CDH의 통합: 최적의 ecosystem 프로젝트의 통합 버전
- · 고 성능 SQL 분석 기능 제공(기존 버전에 대한 성능 개선) : Impala, Tez
- · 실시간 스트리밍 처리 및 분석 기능 제공 : Kudu
- 세분화된 보안 및 데이터 거버넌스 기능 제공 : Ranger, Atlas + Navigator
- ON PREM환경에서의 Object Storage 지원 : Ozone(Tech Preview)
- · 클러스터 관리 기능 강화 : Cloudera Manager

CDP Data Center 주요 기능 상세

기존 CDH 고객에게 제공되는 기능 혜택	
Ranger	 동적 Row Filtering, Column 마스킹 사용자 및 속성(Attribute) 기반의 접근 제어 SparkSQL에 대한 세분화된 접근 제어
Atlas	개선된 데이터 탐색 기능 제공성능 및 확장성 개선
Hive 3	 Hive on Tez를 활용한 ETL 성능 개선 ACID Transaction 지원
Knox (*)	• 게이트웨이 기반의 SSO 기능 제공
Druid (*)	• 실시간 및 집계성 데이터에 대한 Data Mart
Spark on Docker(*)	• 단순화된 배포 관리

기존 HDP 고객에게 제공되는 기능 혜택	
Cloudera Manager	 다중 클러스터를 단일 도구로 관리 쉽고 편한 보안 클러스터 설정 지원 CM 관리자에 대한 세분화된 접근 제어 지원(RB AC) 외 다양한 관리 기능 제공
Navigator	• 개선된 Data Lineage 기능 제공
Impala	 Data Mart 워크로드를 마이그레이션을 위한 고성능 SQL 엔진 지원 HUE SQL Editor를 통한 Interactive BI 업무 수행
Kudu	• 변경되는 데이터에 대한 실시간 분석 지원
Data Encry ption	Key Trustee Server, Navigator Encrypt (*)

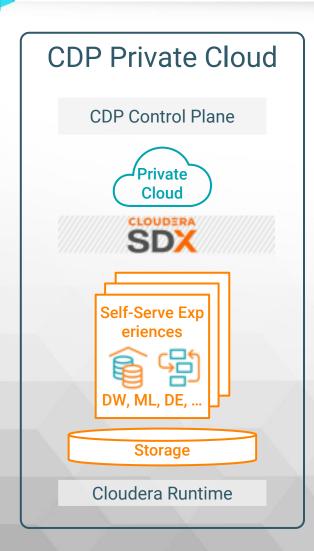
모든 고객에게

- Virtual Private Clusters : Compute & Storage 계층을 분리하여 활용
- Ozone Object Storage (Tech Preview): HDFS 대비 10배 확장성 제공

* Roadmap

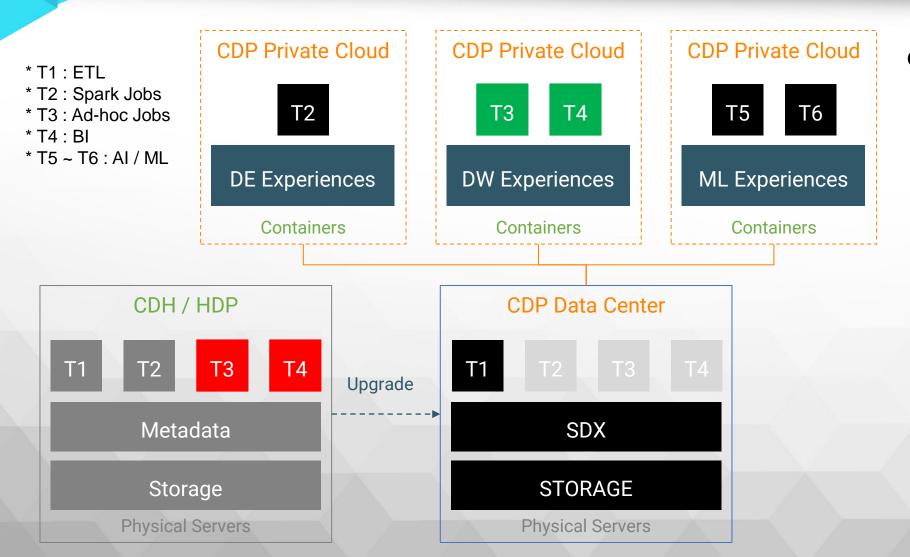


CDP Private Cloud 개요



- 1. CDP Public Cloud와 비교하면
 - · On-Premise 환경에서 실행
 - 고객이 관리
 - · Cloudera Storage에서 실행
 - · Public Cloud와 동일한 Experience 기능 제공(순차 제공 예정)
 - Data Hub Service 사용 불가
- 2. CDP Data Center와 비교하면
 - 프로비저닝, 확장, 관리 및 업그레이드가 쉬움
 - 완벽한 Multi-tenancy 환경 제공 : 물리적으로 격리된 환경
 - 더 향상된 인프라 자원 활용 환경

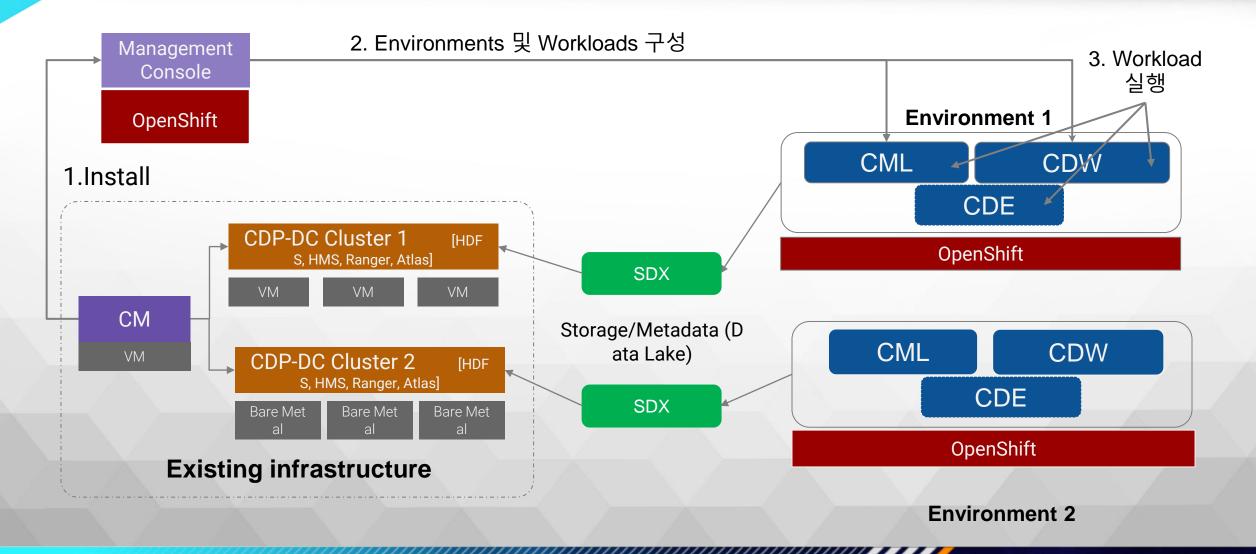
CDP Private Cloud 아키텍처



CDP PVC 혜택

- Storage와 Compute가 분리된 서 비스 제공
- 데이터 공유 : Single source of tr uth
- Public Cloud와 동일한 서비스 제공 : DW / ML Experience
- **격리된 실행 환경 제공** : Noisy N eighborhood 격리
- Sand-Box 서비스 제공 편의성 증 대
- 일관된 사용자 경험 제공: 보안,
 데이터 카탈로그, 데이터 거버넌
 스
- Hardware Utilization 최적화

CDP Private Cloud 구축 방안



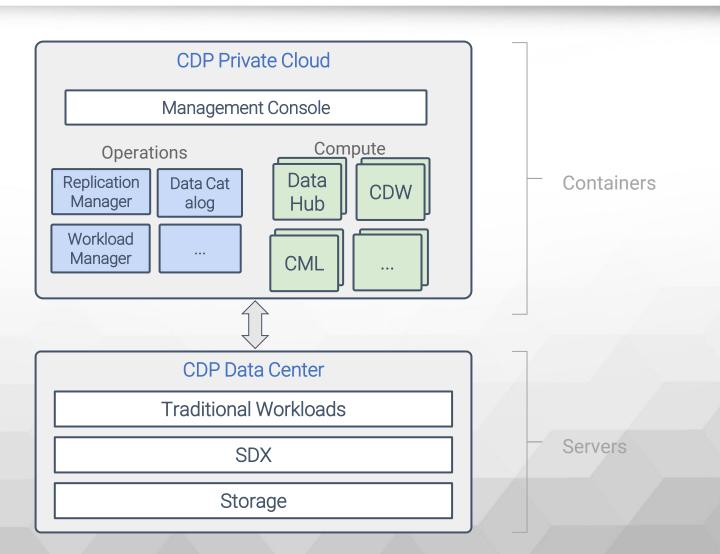
CDP HYBRID CLOUD

CDP DC(Data Center)

· 기존 HDP 및 CDH의 통합된 버전

CDP PVC(Private Cloud)

· CDP Private Cloud를 위해서는 먼저 CDP DC가 구축되어야 합니다.

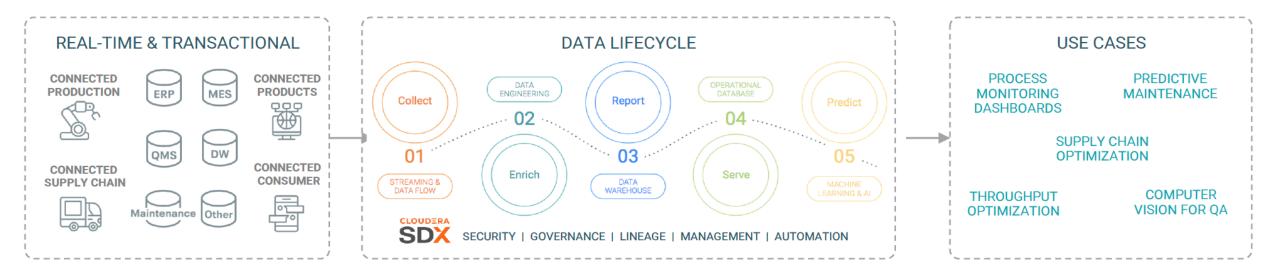


CDP HYBRID CLOUD

비즈니스 사용 사례에는 데이터 수명 주기가 필요합니다

BUSINESS DATA

HYBRID DATA CLOUD INDUSTRY USE CASES



수명 주기가 통합된 하이브리드 데이터 클라우드는 사용, 관리 및 보안이 더 쉽고 빠릅니다.



감사합니다.