CLOUDERA
Educational Services

CDP Private Cloud Fundamentals

CLOUDERA
Educational Services

The Enterprise Data Cloud Vision

The Enterprise Data Cloud Vision

- A new kind of platform

- Four key characteristics

Characteristics of the Enterprise Data Cloud
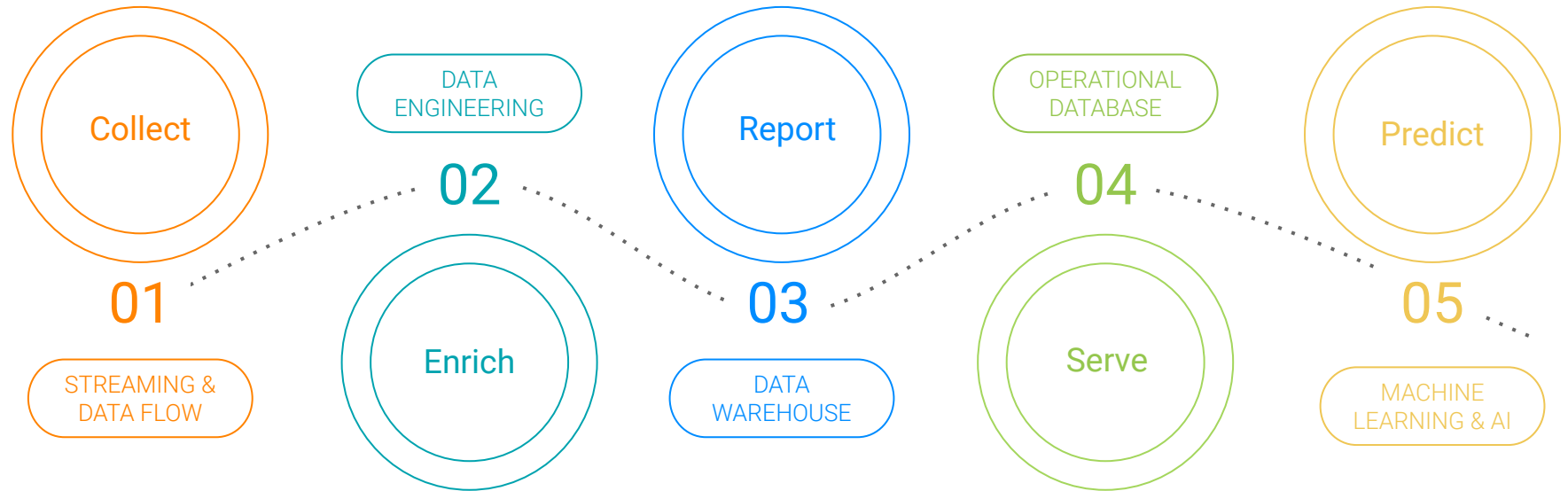

Hybrid & Multi-Cloud

Multi-Function

Secure & Governed

Open

CLOUDERA
Educational Services

Data Lifecycle

Collect

01
STREAMING & DATA FLOW

DATA ENGINEERING
02
Enrich

Report
03
DATA WAREHOUSE

OPERATIONAL DATABASE
04
Serve

Predict
05
MACHINE LEARNING & AI

CLOUDERA SDX SECURITY | GOVERNANCE | LINEAGE | MANAGEMENT | AUTOMATION

CLOUDERA
Educational Services

Cloudera Data Platform Overview

Cloudera Data Platform: Recap

- Control costs and manage resources
  - Auto-scale
  - Auto-suspend

- Easy provisioning and support for multiple types of workloads

- Consistent security and data governance across applications and datasets

- Enables enterprise IT staff to quickly respond to business demands

CLOUDERA
Educational Services

CIO Magazine

How to Eliminate Shadow IT

"Enterprise IT doesn't operate
at the speed of business.

So business users
build their own capabilities
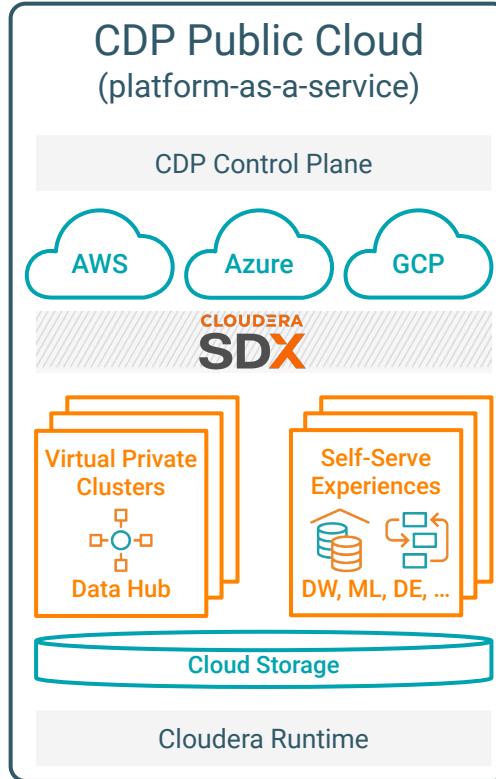through shadow IT purchases."

CIO Magazine

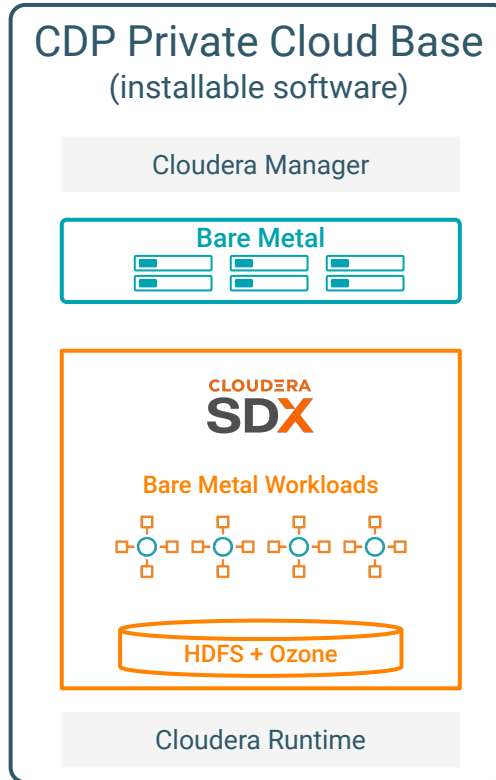"Your IT group needs to perform better than shadow IT."

CIO Magazine

CDP Public Cloud

- Runs on public cloud infrastructure

- Uses cloud provider's object store
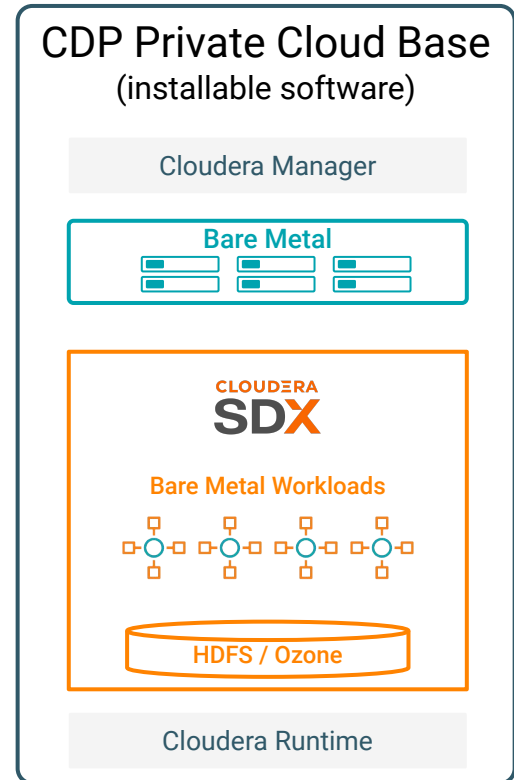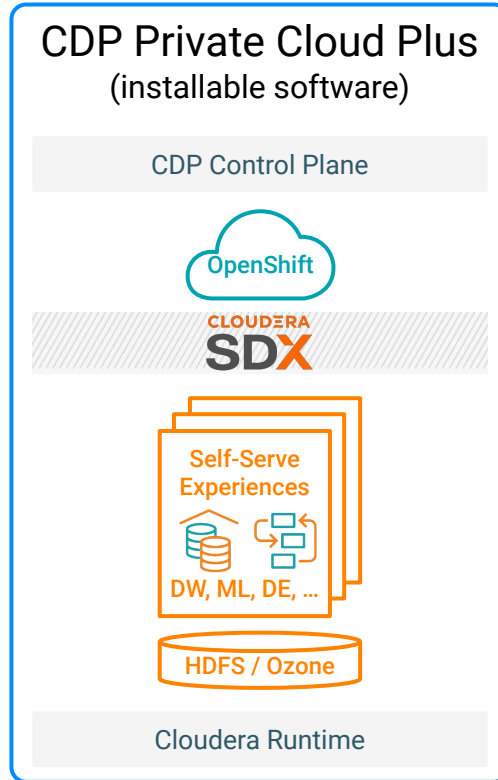
CDP Private Cloud Base

- Similar architecture to CDH and HDP
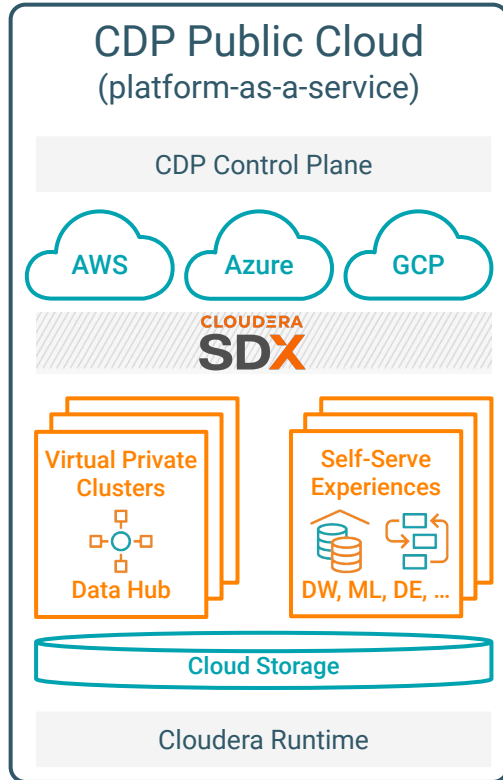
- Formerly known as CDP Data Center

CDP Private Cloud Base
(installable software)

Cloudera Manager

Bare Metal

CLOUDERA
SDX

Bare Metal Workloads

HDFS + Ozone

Cloudera Runtime

CLOUDERA
Educational Services

CDP Product Overview

CDP Public Cloud
(platform-as-a-service)

CDP Control Plane

AWS • Azure • GCP

CLOUDERA SDX

Virtual Private Clusters — Data Hub

Self-Serve Experiences — DW, ML, DE, ...

Cloud Storage

Cloudera Runtime

CDP Private Cloud Plus
(installable software)

CDP Control Plane

OpenShift

CLOUDERA SDX

Self-Serve Experiences — DW, ML, DE, ...

HDFS / Ozone

Cloudera Runtime

CDP Private Cloud Base
(installable software)

Cloudera Manager

Bare Metal

CLOUDERA SDX

Bare Metal Workloads

HDFS / Ozone

Cloudera Runtime

CLOUDERA
Educational Services

CDP Private Cloud Plus

- Runs on private cloud (OpenShift)

- Uses local storage (HDFS / Ozone)

- Depends on a Private Cloud Base cluster
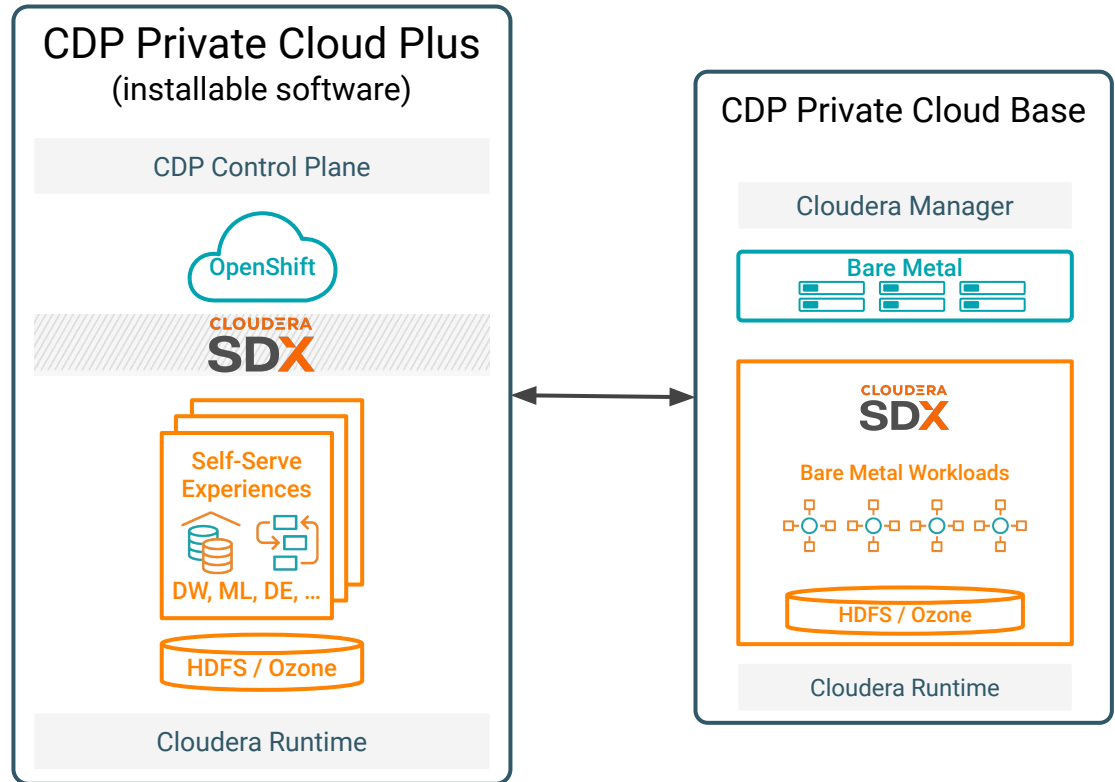
CLOUDERA
Educational Services

CDP Private Cloud Architecture

Important Trends

- Rising cloud adoption increases the pace of business

- This changed user's expectations for IT
  - Instant provisioning
  - Instant scalability

- It has also led to innovations in software architecture

Traditional Cluster Architecture (Bare Metal)

- This cluster might be running
  - CDH
  - HDP
  - CDP Private Cloud **Base**

- Nodes are connected via a network switch
  - Gigabit ethernet was once common

- Designed to conserve limited bandwidth
  - Storage and compute are colocated

- Specific services deployed to specific nodes

This node assigns tasks and tracks where data is stored

The remaining nodes do the actual work of running tasks and storing the data

Limitations of the Traditional Cluster Architecture

- Colocation of storage and compute
  - Can't scale them independently

- Optimized for large files
  - Leads to the "small files" problem

- Shared resource model for multitenancy
  - Leads to "noisy neighbor" problem

- Rigid mapping of services to nodes
 - Distributes resources inefficiently

Key Aspects of the Cloud-Native Architecture

- **Fast networks enable separation of storage from compute**
  - This allows administrators to scale them independently

- **Object stores are the preferred way to store data**
  - This eliminates the "small files" problem

- **Containers decouple an application from the environment where it runs**
  - They provide isolation needed to solve the "noisy neighbor" problem
  - They also enable more efficient distribution of resources

What is Kubernetes?

- Often abbreviated as k8s

- Software system used to deploy, scale, and manage containerized applications

- Originally developed at Google, now open source

- Supported by all major cloud providers and available in commercial products

- A collection of machines running Kubernetes software is called a "cluster"
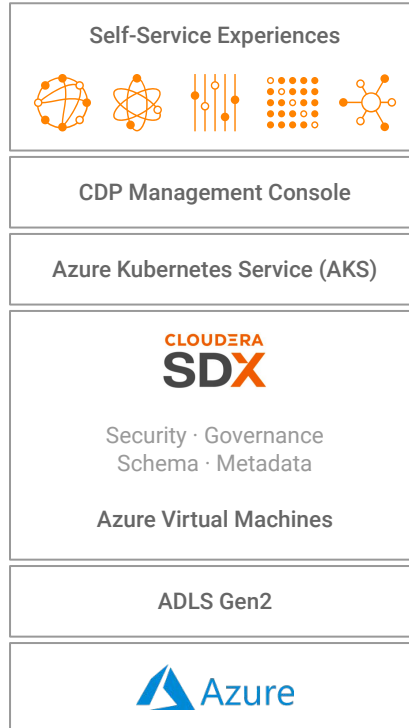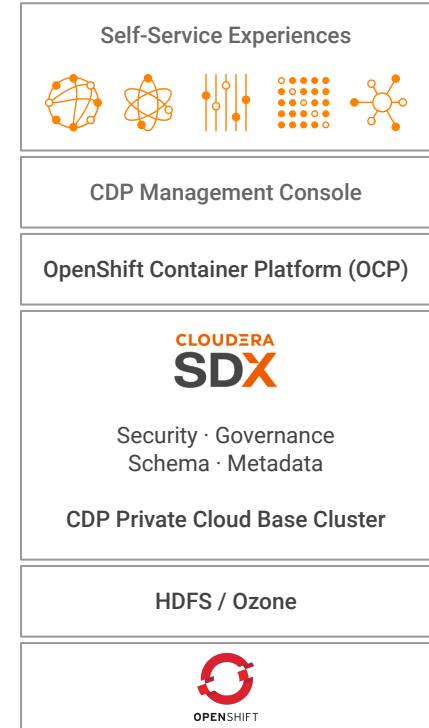
Kubernetes Overview

CLOUDERA
Educational Services
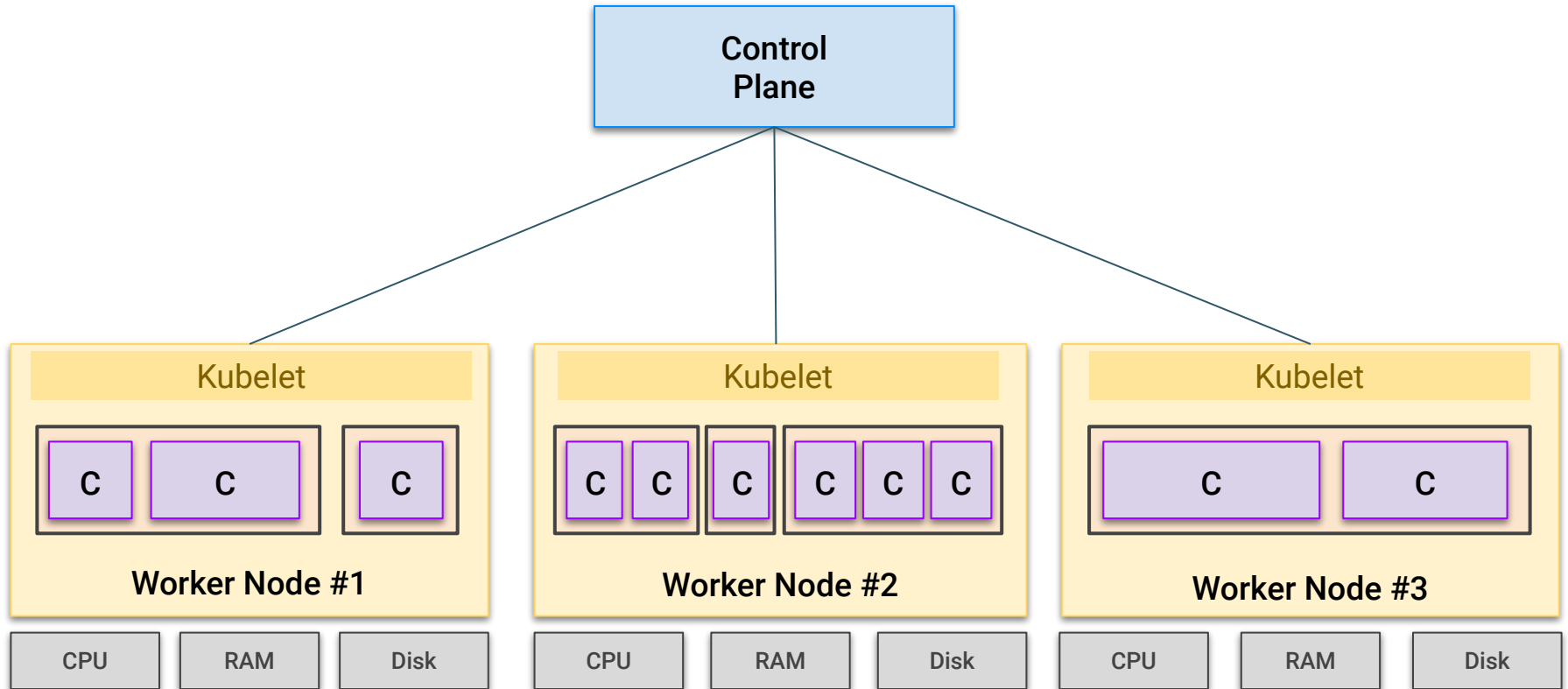
Comparing CDP Public and Private Cloud
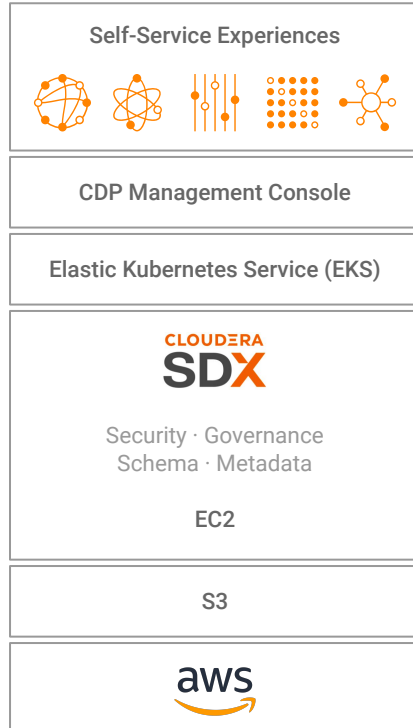
CDP Private Cloud Architecture

CLOUDERA
Educational Services
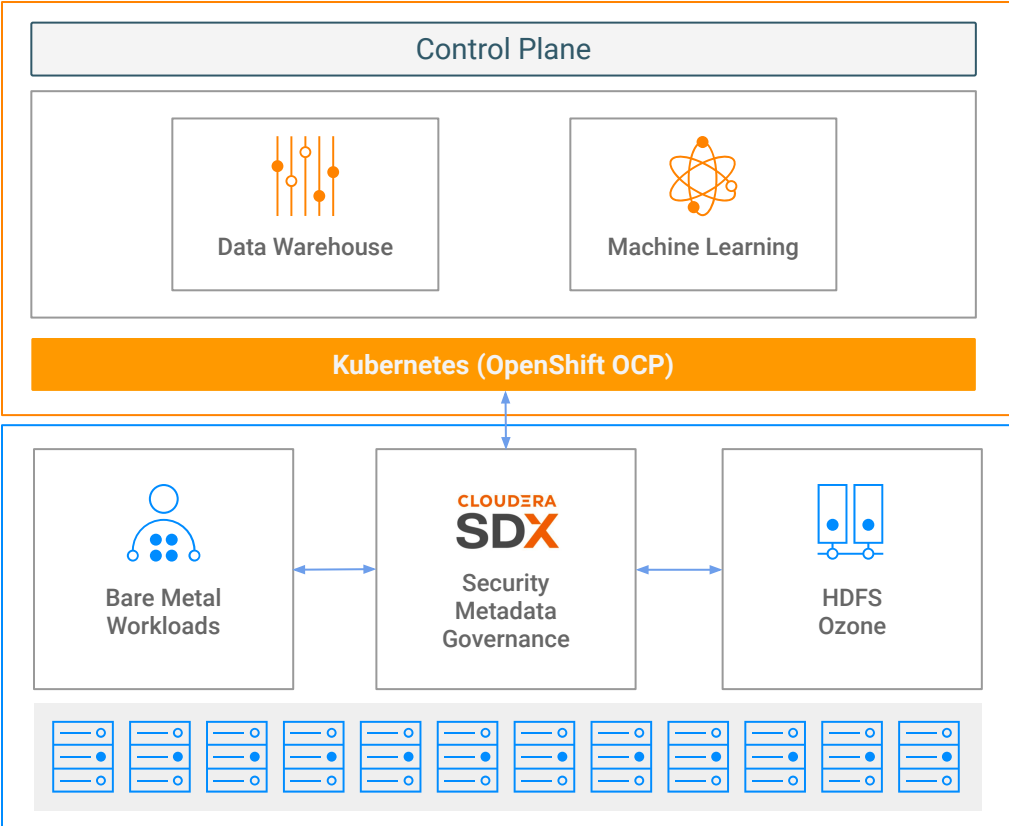
CDP Private Cloud: Initial Release

CLOUDERA
Educational Services
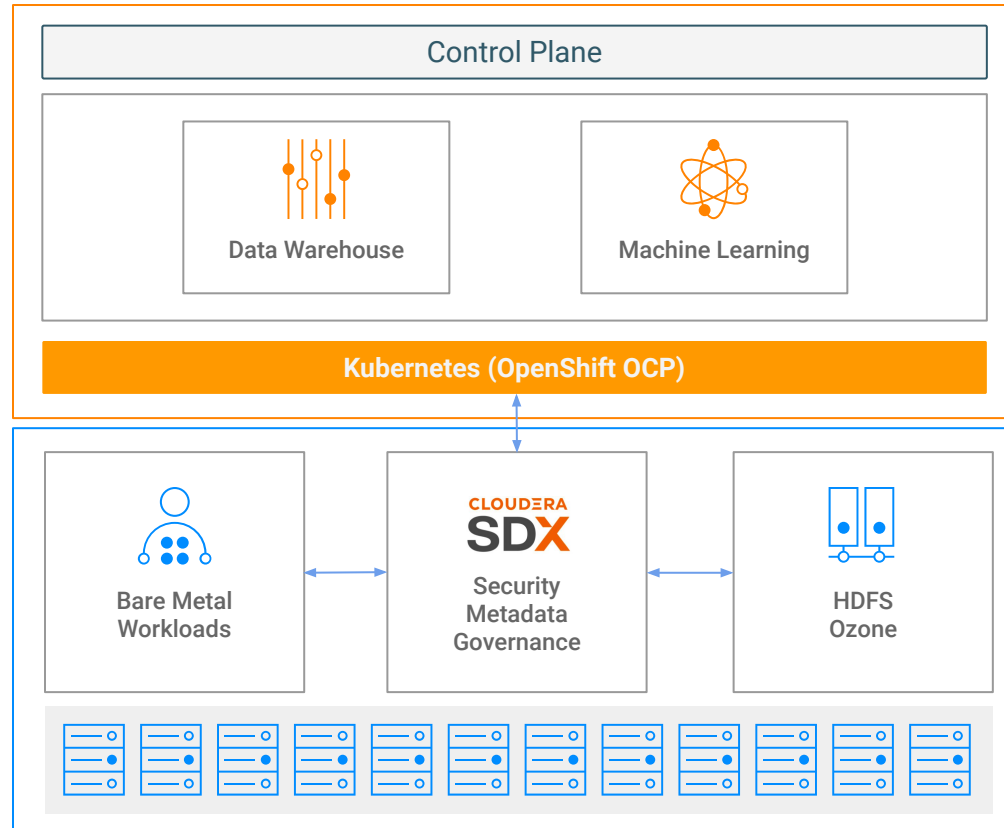
CDP Private Cloud: Future State

CLOUDERA
Educational Services
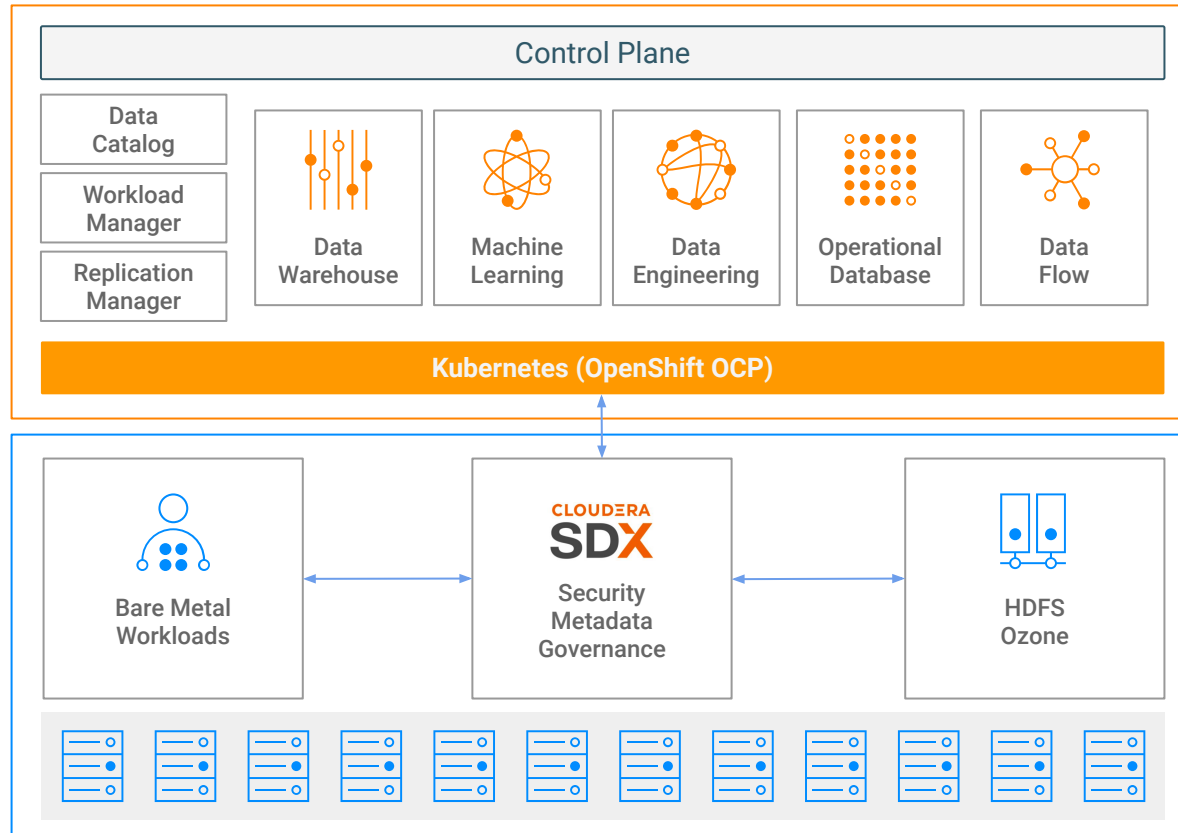
CLOUDERA
Educational Services

Installation

Installation Requirements

- Check the documentation, as details may change

- Red Hat OpenShift Cluster

- CDP Private Cloud Base Edition cluster
  - Serves as the data lake
  - HDFS, Ozone, Hive metastore, Ranger, and Atlas services are required
  - Can upgrade existing CDH/HDP cluster or perform new installation
  - Must be configured for Kerberos (MIT or Active Directory)
  - Must have Auto TLS enabled

- Fast network connection between OpenShift and Data Lake clusters

Installation Demo

- Your instructor will now demonstrate CDP Private Cloud installation

CLOUDERA
Educational Services

Conclusion

Essential Points

- **Simplified multitenancy**
  - Scale the workloads, not individual services

- **Infrastructure agility**
  - Manage shared resources efficiently

- **Upgrade agility**
  - Enable flexibility through containerization

- **Self-service provisioning**
  - Eliminate bottlenecks while maintaining oversight

CLOUDERA
Educational Services