

CDP Form Factors

Cloudera Data Platform (CDP) offers versatility through its various form factors, catering to different business requirements and infrastructure models. The three primary form factors are Public Cloud, Private Cloud, and Data Center.

ONE PLATFORM – TWO FORM FACTORS

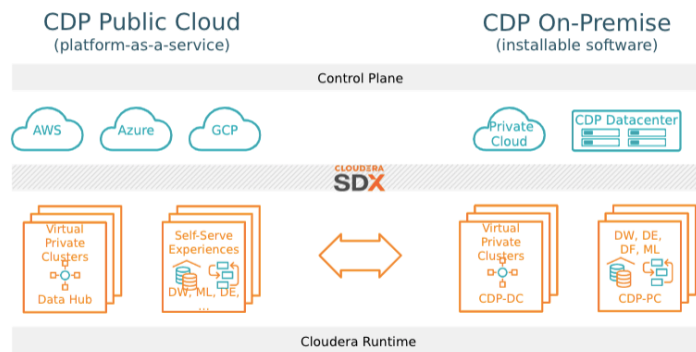


Figure 1: CDP

1. Public Cloud:

- **Architecture:** CDP Public Cloud is designed to leverage the scalability and flexibility of public cloud providers like AWS, Azure, and Google Cloud Platform. It utilizes cloud-native services for storage (like S3, ADLS, or GCS) and integrates with cloud-specific tools and services.
- **Use Case Example:** A retail company uses CDP Public Cloud to process and analyze customer data across various regions. They leverage the scalability of the cloud to handle peak shopping periods without the need for permanent infrastructure investment.
- **Code Snippet (AWS S3 Integration):**

```
# Python snippet to access data stored in AWS S3
import boto3
s3 = boto3.client('s3')
bucket_name = 'your-bucket-name'
object_key = 'data/path/file.csv'
response = s3.get_object(Bucket=bucket_name, Key=object_key)
data = response['Body'].read().decode('utf-8')
```

2. Private Cloud:

- **Architecture:** CDP Private Cloud is designed for on-premises environments but with cloud-like capabilities. It allows businesses to create a private, cloud-native experience in their data centers, offering better control over data and resources.
- **Use Case Example:** A financial institution uses CDP Private Cloud to meet stringent data security and regulatory requirements while benefiting from the elasticity and modern data analytics capabilities of a cloud-native architecture.

3. Data Center:

- **Architecture:** CDP Data Center is the evolution of the traditional on-premises Hadoop deployments. It is suited for businesses that have significant investments in on-premises infrastructure and require complete control over their data environment.
- **Use Case Example:** A healthcare provider uses CDP Data Center to manage sensitive patient data. They require a robust environment that adheres to strict compliance and privacy regulations.

Best Practices for Choosing the Right Form Factor:

1. **Assess Data Sensitivity and Compliance Requirements:** Choose Private Cloud or Data Center for highly sensitive data or strict regulatory environments.
2. **Consider Scalability Needs:** Public Cloud is ideal for businesses that experience fluctuating data processing demands.
3. **Evaluate Existing Infrastructure:** If there's substantial existing on-premises investment, CDP Data Center might be the most cost-effective option.
4. **Analyze Cost Implications:** Public Cloud can offer cost savings for businesses without the need for heavy upfront investment in infrastructure.
5. **Plan for Future Growth:** Consider potential scalability and global reach. Public Cloud provides easy scaling options and global data center access.