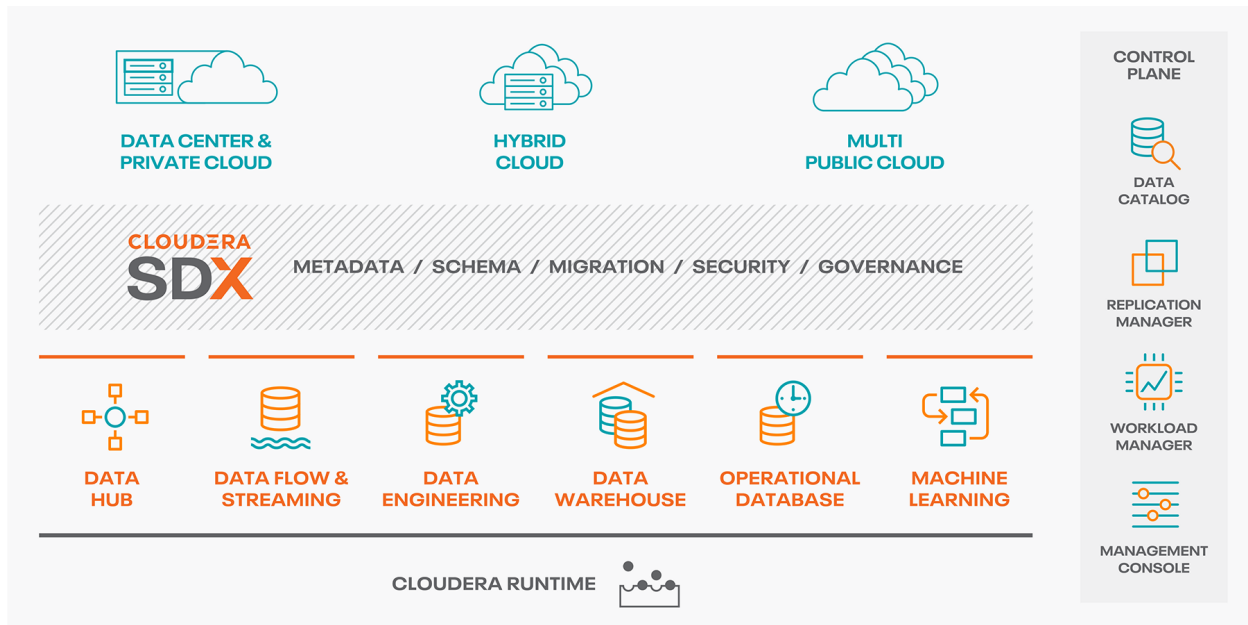


CLUSTERA PLATFORM SERVICE DESCRIPTION

Clustera Platform (CDP)

CDP is an integrated data platform that is easy to deploy, manage, and use. By simplifying operations, CDP reduces the time to onboard new use cases across the organization. It autoscales workloads up and down for more cost-effective use of cloud infrastructure.

CDP manages data in any environment, including multiple public clouds, bare metal, private cloud, and hybrid cloud. With Clustera's Shared Data Experience (SDX), the security and governance capabilities in CDP, IT can confidently deliver secure analytics running against data anywhere. CDP is a new approach to enterprise data, anywhere from the Edge to AI.



SDX brings together the common shared services in CDP. It seamlessly surfaces key platform services, driving productivity and simplicity. The Management Console provides a single pane of glass to create and maintain platform users, roles and access. The Data Lake Service leverages SDX for safe, secure, and governed data lakes wherever data is stored, from object stores to HDFS.

SDX enables the standard enforcement of granular, dynamic, role- and attribute-based security policies. Prevent and audit unauthorized access to sensitive or restricted data across the platform. Encrypt data across the stack at rest as well as in motion and manage encryption keys.

SDX maintains the metadata required to Identify and manage sensitive data and effectively address compliance requirements with unified data management operations, including metadata search, lineage and chain of custody, auditing and security as well as classification, profiling and business glossary.

SDX provides services for Migrating data and associated metadata, complete with security and governance policies, between environments, including legacy clusters, to deliver it there where the enterprise needs to work. Replicate data and metadata to provide backup and disaster recovery services between any environment, from on-premises to hybrid- and multi-cloud.

Cloudera Data Platform - Public Cloud

Cloudera Data Platform - Public Cloud is a big data platform for both IT and the business. Cloudera Data Platform (CDP) is:

- Simple to use and secure by design
- Manual and automated
- Open and extensible
- For data engineers and data scientists

IT has traditionally struggled rolling out bare metal infrastructure, sharing multi-tenant workloads on full-capacity servers, and managing data centers lacking capacity with maxed-out capital budgets. CDP Public Cloud provides welcome flexibility to IT and agility to business users by extending public cloud innovation to the data center, separating compute and storage for better control and agility, delivering a PaaS user experience, and bursting to the cloud when necessary.

CDP Public Cloud provides self-service access to integrated, multi-function analytics on centrally managed and secured business data while deploying a consistent experience anywhere—on premises or in hybrid and multi-cloud. Enjoy consistent data security, governance, lineage, and control, while deploying the powerful, easy-to-use cloud analytics experiences business users require and eliminating their need for shadow IT solutions.

CDP Public Cloud provides a platform for High value, data-driven business use cases that require modern, streaming real-time data and integrated analytics and machine learning service that are both easy for IT to manage and deploy and easy for business users to consume and operationalize. CDP makes it easy to deploy modern, self-service analytics and machine learning services for any data, anywhere, with shared security and governance and the flexibility to scale anywhere with the same experience.

CDP Public Cloud is available on Amazon Web Services (AWS), Microsoft Azure (Azure) and Google Cloud Platform (GCP) enabling a true, hybrid cloud architecture.

Cloudera Data Warehouse

The Cloudera Data Warehouse service enables self-serve creation of independent data warehouses and data marts for teams of business analysts without the overhead of bare metal deployments.

In Cloudera Data Warehouse, your data is stored in an object store in a data lake that resides in your specific cloud environment. The service is composed of:

- Database Catalogs:
A logical collection of metadata definitions for managed data with its associated data context. The data context is comprised of table and view definitions, transient user and workload contexts from the Virtual Warehouse, security permissions, and governance artifacts that support functions such as auditing. One Database Catalog can be queried by multiple Virtual Warehouses.
- Virtual Warehouses:
An instance of compute resources that is equivalent to a cluster. A Virtual Warehouse provides access to the data in tables and views in the data lake that correlates to a specific Database Catalog. Virtual Warehouses separate compute and storage by executing queries on tables and views that are accessible through the Database Catalog that they have been configured to access.

The Cloudera Data Warehouse service provides data warehouses and data marts that are:

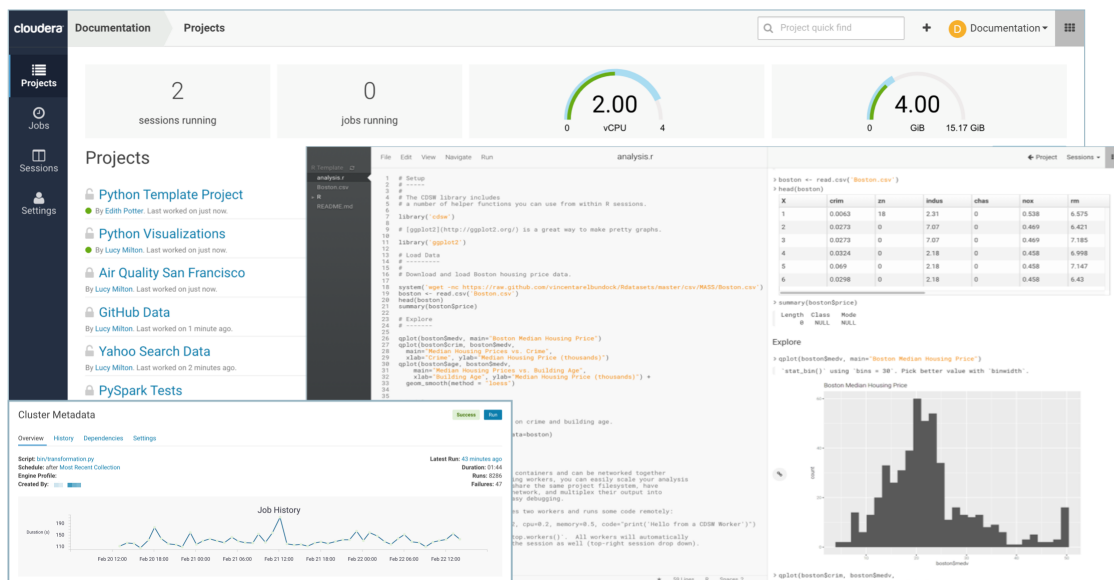
- Automatically configured and isolated
- Optimized for your existing workloads when you move them to the cloud
- Auto-scaled up and down to meet your workloads' varying demands
- Auto-suspended and resumed to allow optimal usage of resources to save costs
- Compliant with the security controls associated with your data lake

<https://docs.cloudera.com/data-warehouse/cloud/index.html>

Cloudera Machine Learning

Cloudera Machine Learning (CML) is Cloudera's cloud-native machine learning service, built for CDP. The CML service provisions clusters, also known as ML workspaces, that run natively on Kubernetes.

Each ML workspace enable teams of data scientists to develop, test, train, and ultimately deploy machine learning models for building predictive applications all on the data under management within the enterprise data cloud. ML workspaces are ephemeral, allowing you to create and delete them on-demand. ML workspaces support fully-containerized execution of Python, R, Scala, and Spark workloads through flexible and extensible engines.



The Cloudera Machine Learning service provides collaborative Workspaces that enable;

- Seamless portability across private cloud, public cloud, and hybrid cloud powered by Kubernetes
- Bring your own editor including Jupyter, RStudio and PyCharm
- Rapid cloud provisioning and autoscaling of Scala, Spark, Python and R workloads
- Fully containerized workloads - including Python, R, and Spark-on-Kubernetes - for scale-out data engineering and machine learning with seamless distributed dependency management
- High performance deep learning with distributed GPU scheduling and training
- Compliant with the security controls associated with your data lake
- Provision of AMPS and model drift tracking

<https://docs.cloudera.com/machine-learning/cloud/index.html>

Cloudera Data Hub

The CDP Data Hub service provides a managed method of deploying and maintaining collections of services referred to as Runtime Clusters. These services are managed using templates. Cloudera Data Hub is the right service to use if you have custom requirements for your cluster. Clusters are built using the CDP runtime and the build process can be fully automated via the CLI. Datahubs can also run a recipe which is used to perform specific tasks on the Data Hub cluster nodes.

The Cloudera Data Hub service provides clusters that are

- Automatically configured and isolated using templates
- Can be customised to a wide range of workloads
- Can be built and destroyed automatically based on workload demands
- Compliant with the security controls associated with your data lake
- Auto-Scaling compute
- Click of a button upgrades

Cloudera Datahub also includes standard templates for deploying the following managed clusters

- Data Engineering : A Data Engineering cluster provides the services required for processing data at volume. These services include Spark 2.4 and Data Analytics Studio
- Data Mart: A Data Mart service enables a federated SQL experience based on Impala
- Real-time Data Mart: As with the Data Mart service, this provides a federated SQL experience, backed by the Apache Kudu data store.
- Operational Database: Provides a HBase on S3 managed service.
- Cloudera Flow Manager (CFM): Provides NiFi and the NiFi Registry services for ingesting, routing and moving data between services.
- Cloudera Streams Messaging (SMM): Provides a Kafka deployment, the Schema Registry and integrated Streams Messaging Manager for operational insights into topics and SLAs.

<https://docs.cloudera.com/data-hub/cloud/index.html>

Data Catalog

Data Catalog is a service within Cloudera Data Platform that enables you to understand, manage, secure, and govern data assets across the enterprise. Data Catalog helps you understand data across multiple clusters and across multiple CDP environments. You can search and locate relevant data of interest based on various parameters. Using Data Catalog, you can understand how data is interpreted for use, how it is created and modified, and how data access is secured and protected.

Data Catalog enables data stewards across the enterprise to work with data assets in the following ways:

- Organize data based on business classifications, purpose, protections needed, etc.
- Catalog and search to locate relevant data of interest (sensitive data, commonly used, high risk data, etc.)
- Automatic data catalog profilers
- Understand what types of sensitive personal data exists and where it is located
- View basic descriptions: schema, classifications (business cataloging), and encodings
- View statistical models and parameters, user annotations, wrangling scripts, view definitions etc.
- Understand how data is created and modified. Visualize upstream lineage and downstream impact, Understand how schema or data evolve, View and understand data supply chain (pipelines, versioning, and evolution)
- Understand how data access is secured and protected, and audit use. Understand who can see which data and metadata (for example, based on business classifications) and under what conditions (security policies, data protection, anonymization)
- View who has accessed what data from a forensic audit or compliance perspective, Visualize access patterns and identify anomalies

<https://docs.cloudera.com/data-catalog/cloud/index.html>

Cloudera Operational Database

An operational database-as-a-service that brings ease of use and flexibility to Apache HBase, CDP Operational Database enables developers to quickly build future-proof applications that are architected to handle data evolution.

CDP Operational Database empowers developers to automate and simplify database management with capabilities like auto-scale, auto-heal, and auto-tune. It is fully integrated with Cloudera Data Platform (CDP), enabling end-to-end visibility and security with SDX as well as seamless integrations with CDP services such as Data Engineering, Machine Learning, and Data Warehouse.

<https://www.cloudera.com/products/operational-db.html>

Cloudera Data Flow Service

Cloudera DataFlow for the Public Cloud is a new streaming experience on Cloudera Data Platform (CDP) that addresses the operational challenges around deployment, scaling, and infrastructure sizing for Apache NiFi flows and offers comprehensive monitoring capabilities.

DataFlow for Public Cloud addresses the following challenges:

- Sharing resources amongst multiple NiFi flows in the same cluster impacts the overall performance
- Underestimating cluster sizes can result in unplanned infrastructure expense
- Scaling up the infrastructure on an as-needed basis can be an operational nightmare
- Monitoring data flow metrics across multiple clusters in a single view is not possible with current tooling

<https://www.cloudera.com/products/cdf/dataflow-for-public-cloud.html>

Workload Manager

The Workload Manager Service provides you with the tools you need to gain an in-depth understanding of your workloads. It is used to Optimize performance and resource utilization across platform workloads. This helps and operational teams maintain SLAs and deliver improvements in useability and resource efficiencies. This drives up utilisation and reduces costs.

Workload Manager provides

- Complete workload visibility for performance tuning and migration
- Usage-based insight for data model optimization
- An understanding of query performance and improvements
- Intuitive self-service insight for help with support platform

<https://docs.cloudera.com/workload-manager/cloud/index.html>

Replication Manager

Cloudera Replication Manager is part of SDX. Replication Manager provides safety and governance in data availability wherever it resides or moves too. Data and workloads can be migrated between infrastructures or can be configured for disaster recovery and fail-over. Replication includes all security and data governance policies.

Replication Manager provides

- Complete backup and disaster recovery
- Easy migration from legacy clusters to cloud deployments
- Hybrid cloud flexibility through continuous synchronization
- Convenient creation of development and test deployments

<https://docs.cloudera.com/replication-manager/cloud/index.html>

Pricing Overview

CDP is charged according to usage/consumption and billed at an hourly usage rate according to the type of instance used, whether AWS, Azure or GCP. CDP pricing is determined by two dimensions; Applicable Rate Schedule and Payment Method.

The Rate Schedule determines the amount that the customer will pay to Cloudera for services consumed under the usage-based licensing model. They provide an hourly rate for each combination of server instance type and subscription level that a customer may select.

The current rate card alongside the available instances are available on the Cloudera pricing guide here; <https://www.cloudera.com/products/pricing/cdp-public-cloud-service-rates.html>

The Payment Method determines how the customer will pay Cloudera for services consumed under the usage-based licensing model. There are two Payment Methods.

Payment Method 1:

Pre-Pay - Under the Pre-Pay Commit model, the customer prepays for a certain volume of credits available during the contract period. As the customer consumes services during the contract period, the appropriate amount is deducted from their credit based on hourly usage. Accounts are metered and reconciled on a periodic basis (monthly by default) and if there is credit remaining after reconciliation, the process repeats in the next period. If an overage occurs, an invoice is sent to the customer and they will continue receiving invoices in a pay-as-you-go manner until they pre-purchase additional credit. The Pre-Pay Commit payment method is recommended by default as it provides access to discount depending on the level and term of pre-pay commit

Payment Method 2:

Pay As You Go - Under the Pay As You Go model, the customer is invoiced at the end of each billing cycle (i.e. monthly) based on usage during that period. Usage is calculated as: (hourly billing rate) x (number of hours per server instance) x (number of server instances). As explained in the Cloudera Rate Card, the hourly billing rate is listed at the intersection between the Instance Type and the Subscription Level in the Rate Card.

CDP pricing does not include the cost of the underlying third-party cloud services.

For more information, please refer to;

Cloudera Pricing

<https://www.cloudera.com/products/pricing.html>

Cloudera hourly rates

<https://www.cloudera.com/legal/terms-and-conditions/cloud-usage-pricing-terms.html>

Service Onboarding / Offboarding

Before you can start using CDP, you must have a supported 3rd party Cloud account in which to create your CDP environment. You must also be familiar with the 3rd party Cloud services and able to create and configure the resources required for CDP.

Please reference our [Quickstart On-boarding Guides](#)

1. Go to the my.cloudera.com and register a user account
2. Under Applications, select CDP
3. In the CDP console, create a CDP Credential
4. Register a CDP Environment
5. Assign user roles
6. Please Reference the online CDP Quickstarts for AWS and Azure

To offboard from CDP

1. Remove any linked Environments in CDP
2. Please contact [Cloudera Sales](#) to have CDP removed from the Cloudera user account.

Customers can create a free trial account to try out CDP and explore how you could quickly provision clusters and run your workloads in the cloud. The CDP trial account gives you 30 days of access to all CDP services.

After your free trial expires, you can contact [Cloudera Sales](#) if you wish to continue using CDP and to get more information about reusing the workloads that you have created.

Service Constraints

Requires an account on Cloudera.com

Requires a 3rd party cloud account

Requires a suitable cross account role to be available.

Service Security

Using Cluster Connectivity Manager (CCM), CDP can communicate with Data Lake and Data Hub workload clusters that are on private subnets. Communication takes place over private IPs without

any inbound network access rules required. CDP requires that these clusters have outbound connections to the Cloudera control plane. Workload clusters initiate an SSH tunnel to the Cloudera control plane, which is then used for all communication thereafter.

For example, CDP can communicate with clusters that are on private subnets with only private IPs without any additional network configuration or setup. However, CDP requires that these clusters have outbound connections configured.

<https://docs.cloudera.com/management-console/cloud/connection-to-private-subnets/topics/mc-ccm-overview.html>

Backup and Disaster Recovery

Customers have full control over their data and backup strategy. Services are included to enable and carry out backup to Cloud Storage services (e.g. S3, ADLS). Users can recover backups themselves.

Service metadata is backed up by Cloudera and this process is fully automated as part of the service. CDP is managed across different data centres and availability zones.

Support

Cloudera can help you configure, optimize, tune, and run CDP for large scale data processing and analysis.

If you are a Cloudera customer, you can:

- Register for an account to create a support ticket at the support site.
- Visit the Cloudera Knowledge Base.

Further information on the support available to customers is available on our website.

Getting Support

<http://support.cloudera.com>

CDP Private Cloud Base

CDP Private Cloud Base edition provides the foundation components for deploying the Storage, SDX and Control Plane services for CDP Private Cloud. This edition can be deployed on bare metal or Virtual servers, based on our published CDP Private Cloud Reference Architecture.

https://docs.cloudera.com/documentation/other/reference-architecture/PDF/cloudera_ref_arch_cdp_dc.pdf

CDP Private Cloud Base edition provides a managed method of deploying and maintaining collections of services referred to as Runtime Clusters. Clusters are built using the CDP runtime and the build process can be fully automated via the Control Plans.

The service provides clusters that are

- Automatically configured and isolated
- Can be customised to a wide range of workloads
- Can be built and destroyed automatically via automation
- Compliant with the security controls associated with your storage services

Pricing Overview

CDP Private Cloud Base is charged on an annual subscription basis, priced per node. The total number of nodes deployed across a cluster will require a licence. Each node deployed also attracts a variable price element for the amount of compute and storage required above the Cloudera "baseline".

The baseline price is based on each node consisting of 16 cores, 128 GB memory and 48TB of storage. This is taken as the total cores, memory and storage across the cluster, divided by the node count. Any excess above this baseline is charged as per the below.

Variable Price, Compute:

Cloudera Compute Units (CCU) are used to calculate the compute variable price.

A CCU allows additional core and memory capacity of 1 core or 8 GB memory over the baseline. CCUs required are calculated on the largest of these amounts.

Variable Price, Storage:

Variable storage is charged for every 1TB above the base allowance of 48TB per node.

To calculate the overall price; Determine the base price by node. Then add the Variable Price Compute and Variable Price Storage.

Cloudera Pricing

<https://www.cloudera.com/products/pricing.html>

Service Onboarding / Offboarding

On procurement of the service, a Cloudera Account Executive will contact the customer to arrange onboarding. This will include delivery of a license key to enable the enterprise features of the Cloudera software and on-boarding of Primary Support Contact(s) from the customer organization.

A defined process exists whereby within 24 hours, an introduction email and license entitlement is provided to the nominated customer contact. A Proactive Customer Onboarding representative will set up a call within one week to arrange the details required to provide our support services.

When the subscription ends (if it is not renewed), the license key will expire and the customer will no longer have access to the enterprise features in the Cloudera software or Cloudera Support. Access to data is retained by the client.

Service Constraints

The total number of nodes deployed across a cluster will require a licence. Each node deployed also attracts a variable price element for the amount of compute and storage required above the Cloudera "baseline".

The baseline price is based on each node consisting of 16 cores, 128 GB memory and 48TB of storage. This is taken as the total cores, memory and storage across the cluster, divided by the node count. Any excess above this baseline is charged.

Service Security

CDP Private Cloud is integrated with the customers Authorisation, authentication and directory services. CDP utilises strong encryption for all data at rest and uses TLS for all data in motion. It provides integration with Enterprise LDAP, AD and Kerberos standards.

Backup and Disaster Recovery

CDP Private Cloud supported Data Replication between other Private Cloud instances or public cloud instances. These shared services support data snapshots and incremental replicas. As part of the deployment the customer defines the required backup and DR architecture. This can

be across data-centers, across cloud vendors or hybrid between private cloud and public cloud instances.

Support

Cloudera can help you configure, optimize, tune, and run CDP for large scale data processing and analysis.

If you are a Cloudera customer, you can:

- Register for an account to create a support ticket at the support site.
- Visit the Cloudera Knowledge Base.

Further information on the support available to customers is available on our website.

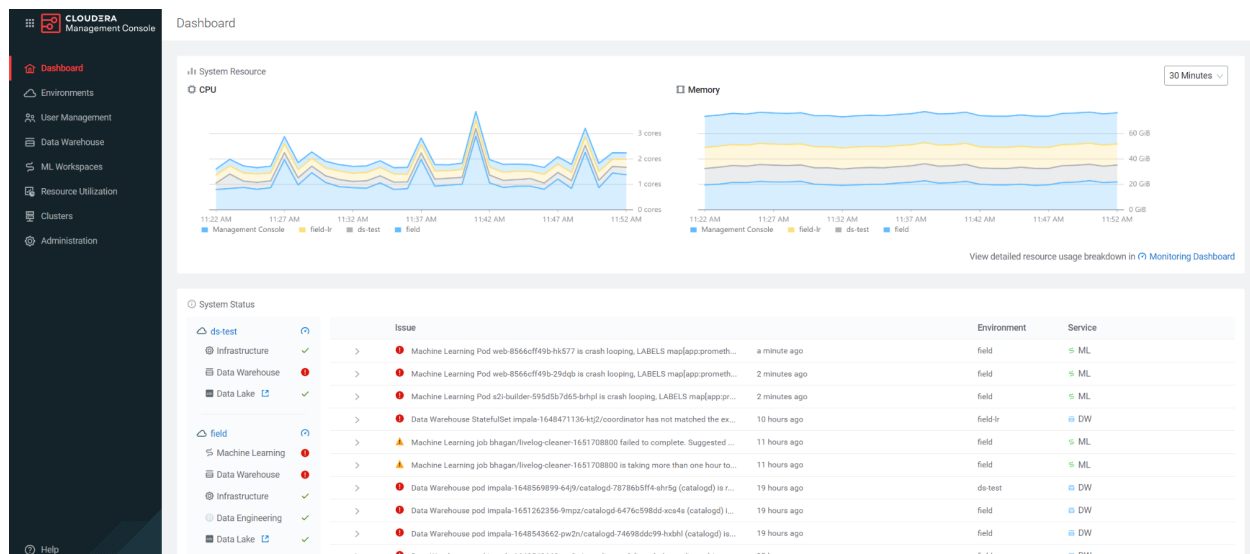
Getting Support

<http://support.cloudera.com>

CDP PRIVATE CLOUD DATA SERVICES

An integral part of CDP Hybrid Cloud, CDP Private Cloud Data Services provides the first step for data center customers toward true data and workload mobility, managed from a single pane of glass and with consistent data security and governance across all clouds, public and private.

- Range of self-service containerized, cloud-native Data Services, delivering faster time to value thanks to simplified on-boarding
- Fully separated, modern compute layer deployed to bare metal or VMs, increasing resource ROI thanks to more efficient utilization
- Private Cloud Base is leveraged for SDX and data storage, reducing risk and improving efficiency with shared, governed data lakes
- Stepping stone to CDP Hybrid Cloud for ultimate flexibility in data and analytics deployment across private and public clouds
- Flexibility to deploy on Openshift or Cloudera provided ECS (Experience Compute Service)



With CDP Private Cloud, organizations benefit from:

- **Rapid time to value** through simplified provisioning of easy-to-use, self-service analytics in minutes rather than days, enabling rapid onboarding of new use cases
- **Improved cost efficiency** with optimized resource utilization and the decoupling of compute and storage, lowering data center infrastructure costs up to 50%
- **Predictable performance** thanks to workload isolation and perfectly managed multi-tenancy, eliminating the impact of spikes on critical workloads and resulting missed SLAs and SLOs

Pricing Overview

CDP Private Cloud Data Services is licensed per Cloudera Compute Unit (CCU) and Cloudera GPU Unit (CGU).

A CCU consists of 1 Core and 8 GBs of RAM. CGUs are based on the type of GPU and number of GPUs running on a given cluster.

Cloudera Pricing

<https://www.cloudera.com/products/pricing.html>

Cloudera hourly rates

<https://www.cloudera.com/legal/terms-and-conditions/cloud-usage-pricing-terms.html>

© 2025 Cloudera, Inc. All rights reserved. [Terms & Conditions](#) | [Privacy Policy and Data Policy](#)

[Apache Hadoop](#) and associated open source project names are trademarks of the [Apache Software Foundation](#). For a complete list of trademarks, [click here](#).

Service Onboarding / Offboarding

On procurement of the service, a Cloudera Account Executive will contact the customer to arrange onboarding. This will include delivery of a license key to enable the enterprise features of the Cloudera software and on-boarding of Primary Support Contact(s) from the customer organization.

A defined process exists whereby within 24 hours, an introduction email and license entitlement is provided to the nominated customer contact. A Proactive Customer Onboarding representative will set up a call within one week to arrange the details required to provide our support services.

When the subscription ends (if it is not renewed), the license key will expire and the customer will no longer have access to the enterprise features in the Cloudera software or Cloudera Support. Access to data is retained by the client.

Service Constraints

CDP Private Cloud depends on a supported version of a Kubernetes container service. Current support includes Red Hat Openshift and Cloudera's Embedded Container Services (ECS). Please review the installation dependencies for more information.

CDP Private Cloud does require the deployment of a CDP Private Cloud Base instance to manage the control-plane and shared services (SDX).

Service Security

CDP Private Cloud is integrated with the customers Authorisation, authentication and directory services. CDP utilises strong encryption for all data at rest and uses TLS for all data in motion. It provides integration with Enterprise LDAP, AD and Kerberos standards.

Backup and Disaster Recovery

CDP Private Cloud supported Data Replication between other Private Cloud instances or public cloud instances. These shared services support data snapshots and incremental replicas. As part of the deployment the customer defines the required backup and DR architecture. This can be across data-centers, across cloud vendors or hybrid between private cloud and public cloud instances.

Support

Cloudera can help you configure, optimize, tune, and run CDP for large scale data processing and analysis.

If you are a Cloudera customer, you can:

- Register for an account to create a support ticket at the support site.
- Visit the Cloudera Knowledge Base.

Further information on the support available to customers is available on our website.

Resources

Public Documentation for Cloudera Data Platform

<https://docs.cloudera.com/cdp/latest/overview/topics/cdp-introduction.html>

Public Pricing information

<https://www.cloudera.com/products/pricing.html>

Cloudera features and pricing

<https://www.cloudera.com/products/pricing/product-features.html>

Getting Support

<http://support.cloudera.com>

Getting Training

<http://university.cloudera.com>