Cloudera Data Platform with IBM

Manage and secure the data lifecycle from the edge to AI in any cloud or data center

IBM

Highlights

— A hybrid data cloud working
  seamlessly across on-
  premises, private, public
  and multi-cloud infrastructures

— Integration of multiple
  analytic tools across
  the data lifecycle

— Enterprise security and
  governance with control
  across all environments

— 100% open-source-based
  software

Cloudera Data Platform (CDP) is a data cloud built for the enterprise. With CDP, businesses manage and secure the end-to-end data lifecycle—collecting, enriching, analyzing, experimenting and predicting with their data—to drive actionable insights and data-driven decisions. The most valuable and transformative business use cases require multi-stage analytic pipelines to process enterprise data sets. CDP empowers businesses to unlock value from large-scale, complex, distributed, and rapidly changing data and compete in the age of digital transformation.

CDP is an integrated data platform that is easy to use for businesspeople and easy for enterprise IT to deploy, manage and secure. It enables people to be more productive using the data and analytics tools they prefer and businesses to operate more efficiently and be prepared to react quickly as business requirements change. It is:

— Designed for data engineers, data scientists, BI analysts,
  developers and enterprise IT
— A simple to use cloud-native service and automatically
  secure by design
— Capable of providing multi-function analytics
  and an integrated data lifecycle
— Self-serve and provides custom analytics
— Able to run seamlessly on public cloud and on-premises

CDP enables enterprise IT to embrace these seemingly opposing options because it delivers the capabilities of an enterprise data cloud.

Cloudera products are used by:

8 of the top 10 global banks

All top 10 global telcos

All top 10 global auto manufacturers

9 of the 10 global pharma companies

8 of the 10 global technology companies[1]

Any cloud: public, private and hybrid

Optimized for hybrid cloud, CDP delivers the same data management and analytic capabilities seamlessly across private and public clouds. Companies can provision a data management and analytic solution in the environment of their choice. CDP separates data management from infrastructure strategy, enabling companies to move data and apps from one environment to another without rewriting applications and retraining personnel. These environments have varying degrees of scalability, flexibility, and cost efficiency.

CDP helps ensure that businesses are able to choose the environment that makes the most sense for their use cases (e.g. data privacy regulations and types of analytics required), business objectives (e.g. speed of deployment and cost efficiency) and technical considerations (e.g. source and location of data and required elasticity).

Data lifecycle integration: From the edge to AI

The integration of streaming, analytics and machine learning—the data lifecycle—is now recognized as a prerequisite for nearly every data-driven business use case. With CDP, companies are able to connect disparate workloads to develop critical apps on one data management platform to address current and future workload needs. By enabling companies to get control of data across all environments, companies are able to master the data lifecycle to get insights that improve productivity and continue their transformation to being data-driven organizations.

Secure and governed

Cloudera's Shared Data Experience (SDX) offers consistent data security, governance, and control across the data lifecycle and across all environments while mitigating risk and costs. It provides data access policies and state information (e.g. metadata, data lineage, metrics and audit trails) as an always-on layer in CDP to meet the needs for ever-increasing regulatory compliance, reduce business and security risks in handling sensitive data, and provide safe and agile self-service access to data and analytics. With SDX, enterprises set data access controls and policies once and they are automatically and seamlessly enforced across data and analytics in hybrid as well as multi-cloud deployments, even as data and workloads move between them.

100% open

The foundation of CDP is composed of open source software (Cloudera Runtime) under either Apache or AGPL licenses. This approach delivers enterprise-grade software that fosters innovation and prevents vendor lock-in. This open source foundation enables CDP to be interoperable with a broad range of analytic and business intelligence applications. The interoperability helps minimize the expense and effort required to connect customers' IT infrastructures with CDP's data and processing capabilities. With CDP, customers can preserve their investment in existing IT architecture as they deploy, integrate and work with unprecedented volumes of structured and unstructured data.

Features of Cloudera Data Platform

There are a number of notable features within Cloudera Data Platform:

Data flow and streaming
Ingest, curate, and analyze data for key insights and immediate actionable intelligence using a highly scalable, real-time streaming analytics solution.

Data engineering
Enrich, transform and cleanse data, making it easy to create, execute and manage end-to-end data pipelines.

Data warehouse
Ingest large-scale data anywhere from structured, unstructured and edge sources into an auto-scaling, highly concurrent and cost-effective analytics service.

Operational database
Serve traditional structured data alongside unstructured data from a real-time, scalable database within a unified open source platform.

Machine learning and AI
Provide enterprise data science teams access to business data and the tools and computing resources required for end-to-end machine learning workflows.

Data hub
Run a wide range of custom analytics with a suite of integrated open-source technologies in a traditional cluster-based environment.

Control plane
Easily monitor, orchestrate and administer data in hybrid and multi-cloud environments through a centralized dashboard.

Shared data experience
Provide comprehensive data privacy, compliance and regulation across the data lifecycle and all clouds.

Why CDP?

CDP is a data cloud built for the enterprise. With CDP, your businesses can better manage and secure the end-to-end data lifecycle—collecting, enriching, analyzing, experimenting and predicting with their data, to drive actionable insights and data-driven decisions. Three of the main reasons enterprises choose CDP are:

Lower total cost of ownership
CDP offers an integrated data lifecycle, eliminating ongoing orchestration and integration development work required to implement business use cases with individual data or analytics services from cloud providers.

Stronger security
CDP SDX provides comprehensive security functions across the data lifecycle to secure data and keep it secure. SDX incorporated identity management, DNS, kerberos, TLS, web proxies, storage encryption, SSO, cloud storage authorization, audit, lineage, data stewardship, context preserved replication and workload optimization. Other vendors provide a subset of these security capabilities and an even more limited set of integration across the lifecycle.

Run anywhere
CDP runs analytic workloads on multiple public clouds, on-premises private clouds or a combination of each for hybrid cloud with a consistent user and IT administration experience. Single cloud providers can't match this flexibility, which is critical for data governance, performance and geographic coverage.

Why IBM for CDP?

IBM builds data management products with the AI, hybrid and multi-cloud future in mind. Its enterprise-grade solutions are designed for robust integration across security-rich environments. To assist with your data lake needs, IBM provides value-added offerings including products, services and multi-vendor support.

For more information

To learn more about Cloudera Data Platform, please contact your IBM representative or IBM Business Partner, or visit ibm.com/analytics/partners/cloudera.

1   According to Cloudera brochure, "We deliver an enterprise data cloud for any data, anywhere, from the Edge to Al" from 2019.