



INDUSTRY

Big data applications and integration

WEBSITE

www.cask.co

COMPANY OVERVIEW

Cask makes building and running big data solutions on-premises or in the cloud easy with Cask Data Application Platform (CDAP), the first unified integration platform for big data. CDAP reduces the time to production for data hubs and data applications by 80%, empowering the business to make better decisions faster.

PRODUCT OVERVIEW

Cask Data Application Platform (CDAP) is an open source, Hadoop-native, enterprise-ready platform for developers, data engineers and citizen integrators to build, deploy, and manage analytics solutions on EDH.

SOLUTION HIGHLIGHTS

- CDAP offers self-service, drag & drop data integration that goes beyond ingest of streaming and batch data from varied sources. It also adds data preparation and data science (ML) for advanced data pipelines on CDH Hadoop and Spark.
- CDAP provides runtime services to enable the complete application life cycle for big data solutions, from development to deployment and managed operations.
- CDAP runs natively on Hadoop for seamless out-of-box scalability over CDH.
- CDAP provides application-level data discovery with metadata, audit and lineage. It seamlessly integrates with Cloudera Navigator.
- CDAP integrates seamlessly with Cloudera Sentry for authentication and fine grained access control.

Cloudera and Cask: From Data Ingest to Machine Learning to Operational Analytics Solutions

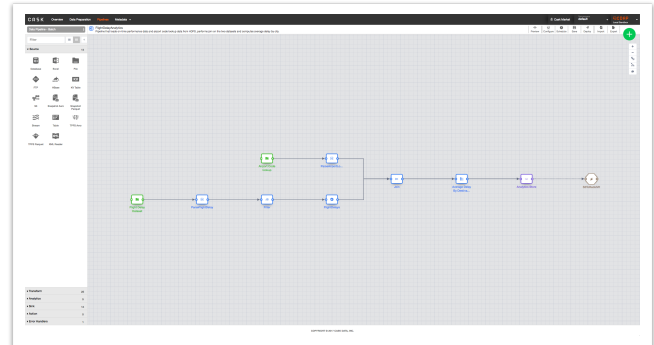
Savvy organizations recognize that data only has value when it’s used to achieve a strategic business objective. Across lines of business, the insights generated from processing, exploring, and analyzing virtually limitless data are changing the way companies compete, bring new products to market, and better service their customers.

Today enterprises are turning to an enterprise data hub (EDH), powered by Apache Hadoop and Spark as the core platform for delivering new analytic applications. Now Cask and Cloudera are bringing together the combined power of large scale processing, data integration and application lifecycle management to make it easier for developers and organizations to build, deploy and operate powerful operational analytics solutions on the EDH.

Delivering Managed Data Lakes and Data Applications

Cask is empowering enterprises to develop and deploy modern analytics solutions by providing the de-facto application and integration layer that abstracts the complexities of big data and allows developers to use familiar skills. With Cask Data Application Platform (CDAP), enterprises can:

- Derive greater value from data by allowing developers and data scientists to focus on finished solutions rather than underlying infrastructure problems.
- Accelerate and simplify application design and development utilizing existing talent and processes.
- Manage the solution life cycle, using enterprise best practices for development, debugging, deployment, and maintenance.



CDAP is the industry’s first unified integration platform for big data that makes it possible to seamlessly apply different technologies to multiple data stores without getting deep into Hadoop. CDAP provides integrated capabilities to build Data-as-a-Service, Data-Ingestion-as-a-Service, and other analytics applications from functional elements in Hadoop and Spark. By integrating and abstracting the complexities of the big data ecosystem, CDAP provides higher order capabilities that more closely mirror the needs of actual use-case implementations. CDAP is native to EDH and runs all of its components within the cluster making it easier to deploy and secure. Additionally, CDAP provides runtime services that enable development, debugging, deployment, and management, all consistent with many IT best practices.

An effective application development, integration and deployment framework depends on a robust platform foundation. Certified by Cloudera and tightly integrated with the EDH, CDAP offers complete application life cycle management on Hadoop and Spark together with an enhanced, intuitive user experience designed for modern self-service environments.

Cloudera Enterprise Benefits

Stores and Analyzes Any Type of Data

- Leverage the full power of your data to achieve pervasive analytics, increase business visibility, and reduce costs
- Bring diverse users and application workloads to a single, unified pool of data on common infrastructure; no data movement required

Enterprise Approach

- Compliance-ready perimeter security, authentication, granular authorization, and data protection through encryption and key management
- Enterprise-grade data auditing, data lineage, and data discovery

Industry-Leading Management and Support

- Best-in-class holistic interface that provides end-to-end system management
- Open platform ensures easy integration with existing systems
- Open source to achieve stability, continuous innovation, and portability

Benefits of CDAP

Accelerate and Simplify Development

- Utilize existing talent and development processes, and use enterprise best practices for development, debugging, deployment and maintenance

Self-Service with Guardrails

- CDAP enables IT to create a self-service experience for data ingestion to data delivery with minimal intervention, while putting in the necessary “guardrails” for enterprise oversight and control

Enterprise-Ready

- CDAP is an open and standards-based architecture that provides extensive security, compliance and resiliency features to support the scale and risk profile of a modern enterprise big data platform

Benefits of CDAP on an EDH

Versatile and Open Source

To maximize innovation in applications, developers need to access a variety of programming options and data stores. An EDH offers the core Hadoop elements of MapReduce, HDFS, and integrated Spark, search and interactive SQL via Cloudera Impala, paired with system management and security capabilities. 100% open source and extensible, CDAP accelerates the design and deployment of big data analytics by up to 5 times compared to hand-coded systems.

Easy to Use and Self-Service

Utilizing CDAP on an EDH is a seamless experience. CDAP is integrated with Cloudera Manager, enabling customers to install, update, and monitor CDAP directly within the Cloudera Manager user interface. CDAP provides automation for ingestion and exploration of data in Cloudera Impala.

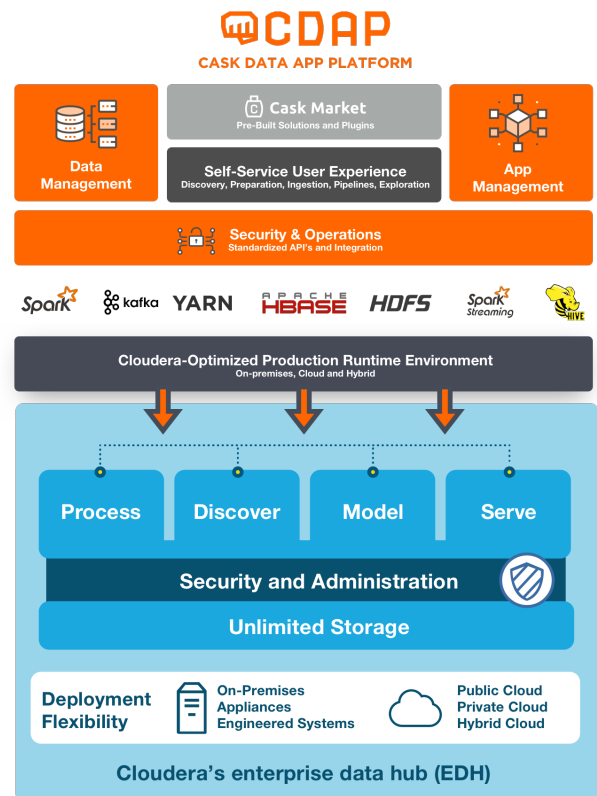
CDAP delivers seamless self-service for transforming, aggregating and enriching large scale batch and real-time data with minimal support from IT. It provides an easy and interactive way to visualize, transform, and cleanse data. It helps data scientists and data engineers derive new schemas and operationalize the data preparation with a few clicks.

CDAP drives faster and more reliable analytics through automatic creation and publishing of datasets, and seamlessly integrates with visualization and data services, making datasets immediately available to reports and applications.

Native to Hadoop and Enterprise-Ready

As organizations move from offline analytics to operational applications, requirements for security, governance, and lifecycle management become critical. The Cloudera EDH includes a variety of embedded security capabilities including authentication, authorization, data governance and lineage, as well as data encryption. CDAP runs natively on Hadoop, offering fault tolerance with automatic recovery along with seamless out-of-box scalability. It provides services for debug and deployment, automated data pipelines, and data discovery. CDAP’s runtime services provide greater transparency at the application level in production, including logging, metrics, and data access lineage. Robust administration features include SLA monitoring, job restart, and error handling, as well as an operations central for auditing access.

Together, CDAP and EDH provide developers and organizations with the confidence that applications can be successfully deployed and managed in production environments.



About Cask

Cask makes building and running big data solutions on-premises or in the cloud easy with Cask Data Application Platform (CDAP), the first unified integration platform for big data. CDAP reduces the time to production for data lakes and data applications by 80%, empowering the business to make better decisions faster. It lets developers, architects and data scientists focus on applications and insights rather than infrastructure and integration. CDAP accelerates time to value from Hadoop through standardized APIs, configurable templates and visual interfaces. It enables IT organizations to broaden the big data user base within the enterprise with a radically simplified developer experience and a code-free self-service environment. CDAP is 100% open source, Hadoop-native and extensible; it seamlessly integrates with existing MDM, BI and security and governance solutions. Cask customers and partners include AT&T, Cloudera, Ericsson, Lotame, Microsoft, Salesforce, and Tableau, among others. For more information, visit the Cask website at cask.co and follow [@caskdata](https://twitter.com/caskdata).

About Cloudera

Cloudera is revolutionizing enterprise data management by offering the first unified Platform for Big Data: The Enterprise Data Hub. Cloudera offers enterprises one place to store, process and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. Founded in 2008, Cloudera was the first and is still today the leading provider and supporter of Hadoop for the enterprise. Cloudera also offers software for business critical data challenges including storage, access, management, analysis, security and search. With over 15,000 individuals trained, Cloudera is a leading educator of data professionals, offering the industry's broadest array of Hadoop training and certification programs. Cloudera works with over 700 hardware, software and services partners to meet customers' big data goals. Leading organizations in every industry run Cloudera in production, including finance, telecommunications, retail, internet, utilities, oil and gas, healthcare, biopharmaceuticals, networking and media, plus top public sector organizations globally. www.cloudera.com.