



Cask

Industry

Big data applications and integration

Website

www.cask.co

Company Overview

Cask provides the de-facto, open source big data application and integration platform that lets developers and data scientists accelerate the time to build, deploy, and operate data-centric solutions and enterprise data hubs. It also enables organizations to implement well-governed data-as-a-service environments designed to quickly unlock the value of data.

Product Overview

The Cask Data Application Platform (CDAP) is an open source, easy-to-use, enterprise-ready platform for developers and organizations to build, deploy, and manage analytics solutions on EDH.

CDAP's self-service extensions, Cask Hydrator and Cask Tracker, offer a drag and drop framework to develop, run, and automate data pipelines on Hadoop and Spark, and to perform data discovery with metadata, audit, and lineage.

Solution Highlights

- CDAP integrates and abstracts multiple technologies such as MapReduce and Spark, and data stores such as HDFS and HBase. It also provides runtime services to enable the complete application lifecycle, from development to deployment and managed operations.
- Cask Hydrator provides self-service, drag and drop data ingestion, and data pipelines on Hadoop and Spark.
- Cask Tracker provides self-service, application-level data discovery with metadata, audit, and lineage.

Cloudera and Cask: Fast Track to Operational Analytics Applications on the Enterprise Data Hub

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, 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.

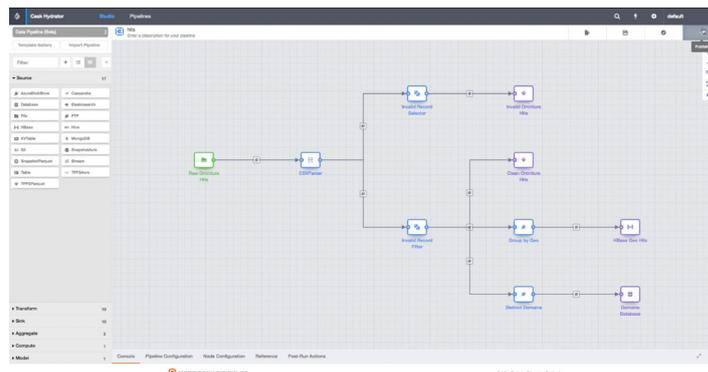
Delivering Data Applications

Cask is empowering enterprises to develop and deploy modern analytics solutions on Hadoop by providing the de-facto application and integration layer that abstracts Hadoop's complexities and allows developers to use familiar skills. With the 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 application and integration layer 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 the complexities of the Hadoop 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. CDAP is certified to work with EDH from Cloudera and offers comprehensive application lifecycle management and an enhanced user experience for Hadoop and Spark.



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

Drive Greater Innovation with Data

- Build a new class of applications that increase business value, including those requiring real-time and batch processing

Accelerate and Simplify Development

- Utilize existing talent and development processes

Manage Your Application Lifecycle

- Use enterprise best practices for development, debugging, deployment and maintenance

Benefits of CDAP on an EDH

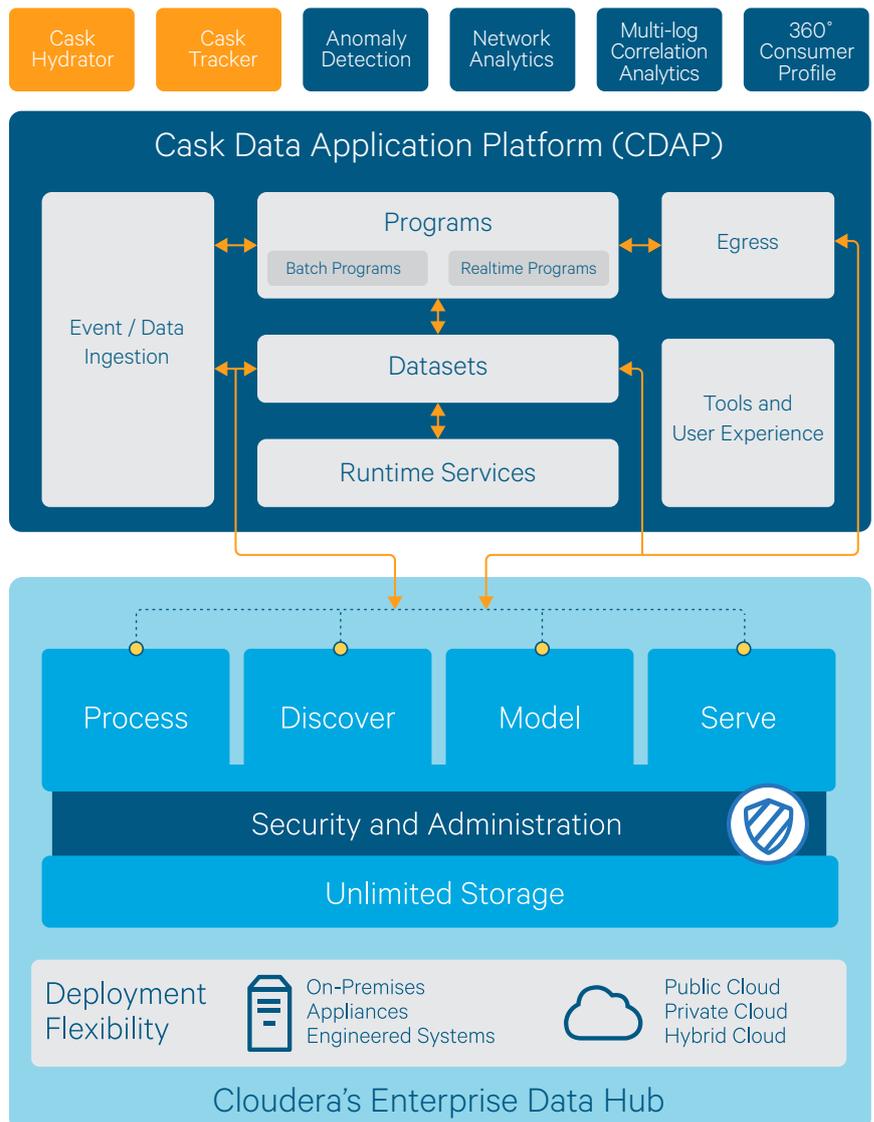
Versatile and Open Source

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, and HDFS, and integrated Spark, search and interactive SQL via Cloudera Impala, paired with system management and security capabilities. As an open source application and integration platform, CDAP makes it easy for developers to utilize the full range of capabilities in an EDH by enabling them to select the programming model they want, without worrying about the underlying data store.

Easy to Use

Utilizing CDAP on an EDH is a seamless experience. CDAP is integrated with Cloudera Manager, enabling customers to install, update, and monitor it directly within the Cloudera Manager user interface. CDAP and Cask Hydrator provide automation for ingestion and exploration of data in Cloudera Impala. Rather than writing a MapReduce program to transform data into an Impala optimized format like Parquet and scheduling periodic transformation jobs in a separate system such as Oozie, developers can issue a few simple commands and either batch or streaming data will automatically be ingested into Impala and made available for high-performance SQL queries.

The alternative is a more manual approach that requires lots of custom scripting, particularly for app development that requires real-time processing.



Enterprise-ready

As organizations move from offline analytics to operational applications, requirements for security, governance, and lifecycle management become critical. An EDH from Cloudera includes a variety of embedded security capabilities, including: authentication, authorization, data governance and lineage, and data encryption. CDAP, with its extensions Cask Hydrator and Cask Tracker, 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. 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 Data ("Cask") provides the de-facto, open source big data application platform that lets developers and data scientists accelerate the time to build, deploy and operate data-centric applications and data lakes/hubs; it also enables IT organizations to implement well-governed data-as-a-service environments designed to quickly unlock the value of data. Through standardized API's, configurable application templates and visual interfaces, Cask's flagship product – the Cask Data Application Platform (CDAP) – cuts down the average implementation time by 80%, while enabling enterprise-class governance, portability, security and transactional consistency in production Hadoop and Spark applications. Major customers and partners include AT&T, Cloudera, Lotame, Pet360, and Salesforce. Cask is headquartered in Palo Alto, California. To download the latest version of CDAP, please visit the Cask website at <http://www.cask.co/>, and to stay updated on product and company news, follow us @caskdata.

About Cloudera

Cloudera delivers the modern data management and analytics platform built on Apache Hadoop and the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems with Cloudera Enterprise, the fastest, easiest and most secure data platform available for the modern world. Our customers efficiently capture, store, process and analyze vast amounts of data, empowering them to use advanced analytics to drive business decisions quickly, flexibly and at lower cost than has been possible before. To ensure our customers are successful, we offer comprehensive support, training and professional services. Learn more at cloudera.com.

cloudera.com

1-888-789-1488 or 1-650-362-0488

Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA

© 2016 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice.