

Cask Data Application Platform

THE UNIFIED INTEGRATION PLATFORM FOR BIG DATA

Harnessing Value from Big Data

The digital transformation of the world and its instrumentation, driven by the web, mobile, and the Internet of Things, are creating massive amounts of data. To harness this data, organizations are investing large amounts of time and money in big data infrastructure such as Apache® Hadoop and Apache® Spark™. However, the low-level and highly disparate nature of these technologies means users have to manually integrate them, resulting in complexity and inefficiencies.

Introducing CDAP

Cask Data Application Platform (CDAP) is the first Unified Integration Platform for Big Data. It provides standardization and deep integrations with diverse open source big data and cloud technologies allowing companies, developers, and users to focus on application logic and insights rather than infrastructure and integration.

The platform is 100% open source, highly customizable, and extensible. It delivers enterprise-class features to help accelerate time to build, deploy, manage, govern, and secure data-driven applications and data lakes on Hadoop and Spark.

App Development

CDAP helps you build, test, and run distributed applications across their entire lifecycle. Simple and easy-to-use APIs help maximize developer productivity, reducing the time to deliver big data solutions. CDAP future-proofs your investment in big data today by seamlessly integrating with new storage and compute technologies as they gain traction in the ecosystem.

Data Integration

CDAP helps you ingest, transform, egress, blend, normalize, index, and track data from diverse sources, types, and formats.

Security & Governance

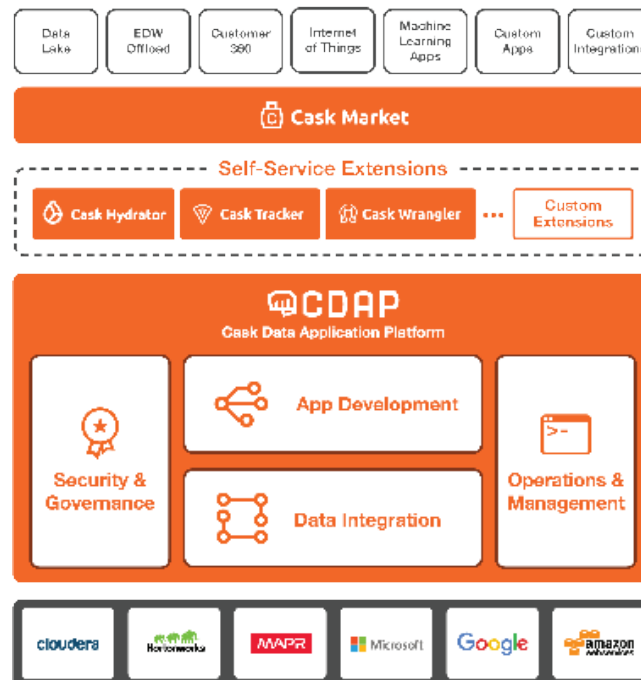
CDAP provides extensive security, compliance, and comprehensive audit capabilities to mitigate risk, improve visibility of data, and identify provenance and lineage.

Operations & Management

CDAP provides a scalable and reliable production runtime environment and operational tools for easy deployment and management of solutions on Hadoop.

Self-Service Extensions

Extensions are rich, intuitive, visual tools that empower users to accomplish tasks with minimal IT support.



CDAP BENEFITS

- Simplify big data integration & app development
- 100% open source, 100% Hadoop native
- Build once, run anywhere
- Java Developers are Hadoop Developers
- Self-service user experience
- Highly extensible

SOLUTIONS

- Data Lake
- EDW Offload
- Customer 360
- Internet of Things
- and more

BIG DATA CERTIFIED

- Cloudera
- Hortonworks
- MapR
- Microsoft Azure HDInsight
- Amazon EMR
- Google Cloud Platform

KEY CUSTOMERS AND PARTNERS

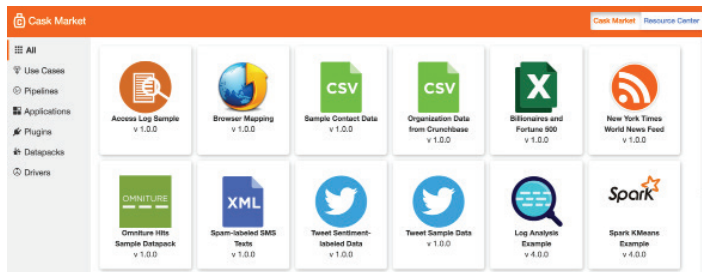
- AT&T
- Salesforce
- Tableau
- Ericsson
- and more

ABOUT CASK

Cask, started by engineers from Facebook and Yahoo, is pioneering the first Unified Integration Platform for Big Data. Based in Palo Alto, Cask is backed by investors such as Andreessen Horowitz, Battery Ventures & Ignition Partners.

Cask Market

Cask Market is Cask’s “Big Data App Store” with push button deployment of pre-built applications, pipelines, and plugins from within CDAP. It provides step-by-step wizards to help configure and deploy new entities within the platform. Companies can easily reuse common components and use cases built internally or published by Cask and its partners.

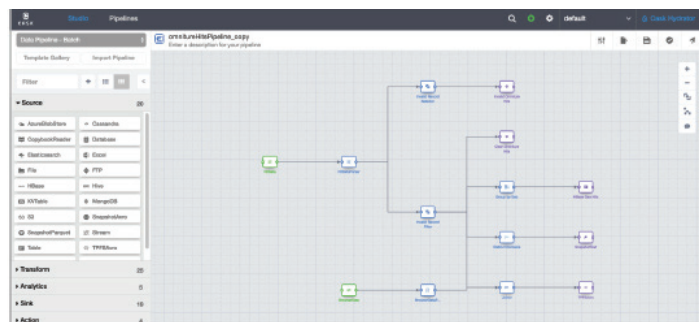


CDAP Extensions

CDAP Extensions are self-service, purpose-built applications on CDAP designed to solve common and critical big data challenges. There are 3 extensions currently packaged with CDAP: Cask Hydrator, Cask Wrangler, and Cask Tracker. Their visual interfaces with rich, pre-built libraries eliminate the need for manual coding and specialized resources, reducing overall complexity. Extensions work seamlessly with one another, streamlining workflows and standardizing processes. All extensions support MapReduce, Spark, and Spark Streaming.

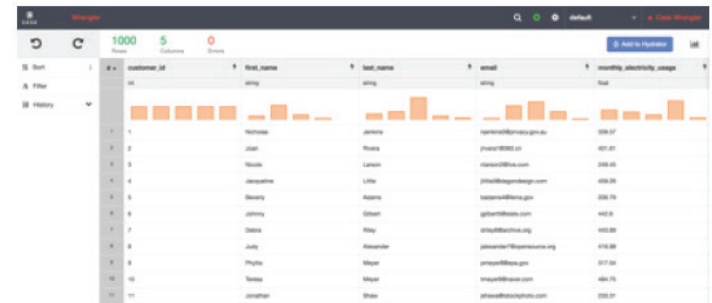
Cask Hydrator

Cask Hydrator is a data ingestion framework that simplifies and automates the difficult and time consuming task of building, running, and managing data pipelines. The studio eliminates coding allowing you to drag-and-drop various sources, transforms, analytics including machine learning algorithms, sinks, and actions, putting big data at your fingertips.



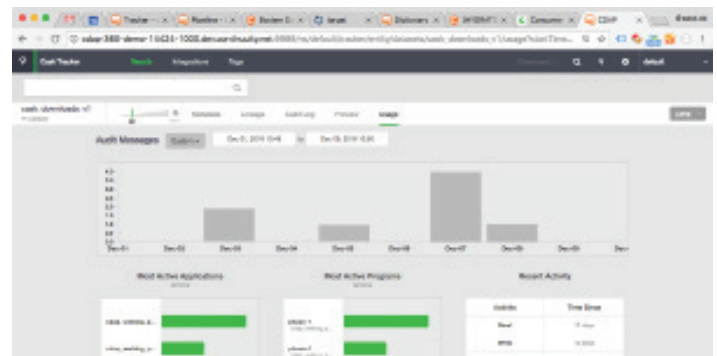
Cask Wrangler

Cask Wrangler provides an easy and interactive way to cleanse, transform, and blend your complex datasets – combined with Cask Hydrator, it makes data transformation not only easier, but faster and more fun, too. Cask Wrangler empowers data scientists and data engineers to inspect, transform, and analyze a variety of data in a self-service fashion, reducing the need for one-off scripts and IT intervention, accelerating time to insights.



Cask Tracker

Cask Tracker helps you discover, profile, and govern data in your data lake. It has powerful features that allow IT as well as business users to manage all facets of data governance including data discovery, metadata tracking, data lineage, and usage analytics. Cask Tracker allows you to perform impact and root-cause analyses, delivering an automated and comprehensive audit trail for compliance. It makes it easy to find, access, understand, and manage datasets.



Summary

CDAP, with its built-in extensions and Cask Market, provides businesses with a single, unified platform that simplifies delivery of a new generation of big data lakes and machine learning apps on Hadoop and Spark. CDAP supports all major Hadoop distros on-premise and in the cloud, as well as being 100% open source and standards driven. Find out more about CDAP at cask.co.

