

Cask Data Application Platform

UNIFIED INTEGRATION FOR BIG DATA

Harnessing Value from Big Data, Cloud and IoT

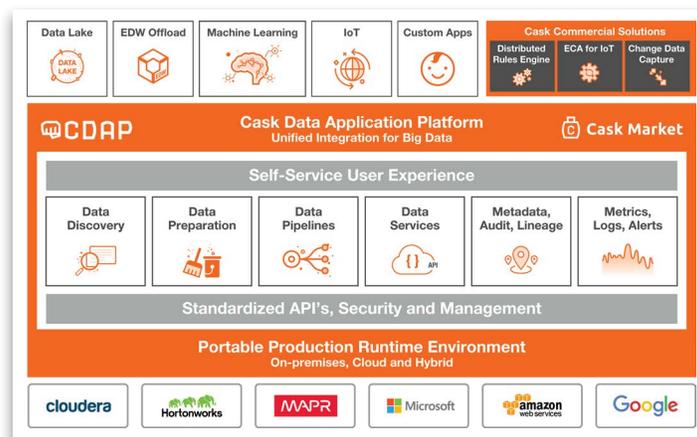
The digital transformation of the world and its instrumentation, driven by the cloud, mobile, and the Internet of Things, are creating massive amounts of data. To harness this data, organizations are investing large amounts of time and money in big data infrastructure such as Apache® Hadoop and Apache®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

From Prep to Production

Cask Data Application Platform (CDAP) is the first Unified Integration Platform for big data. From prep to production, it provides standardization and deep integrations with diverse big data, cloud, and IoT technologies, reducing time to value from your data. CDAP allows developers, data engineers and citizen integrators to focus on application logic and insights rather than infrastructure and integration. CDAP is open source, Hadoop-native, highly extensible and enterprise-ready.

Data and Application Lifecycle Management

CDAP helps you discover, ingest, blend, normalize, index, and wrangle data from diverse sources, types and formats. Offering a unified interface to preview, debug, deploy, run and manage data pipelines, CDAP provides separation of logical data pipelines vs the execution environment, making it easy to run pipelines as MapReduce or Spark jobs. It also provides sophisticated metadata tracking, a data dictionary, and usage analytics to improve visibility, identify lineage, and enforce governance standards. Audit logs offer easy traceability for data quality and compliance needs. CDAP helps you build, test, and run distributed data applications across their entire life cycle. Its simple and easy-to-use APIs help maximize developer productivity, reducing time to value for modern big data solutions.



Portable Production Runtime Environment

CDAP provides a portable production runtime environment and robust security for easy, reliable deployment and management of big data solutions. It offers 'build-once-run-anywhere' flexibility to support infrastructures from on-premises to the cloud and to the edge, as well as hybrid environments.

Self-Service User Experience

CDAP enables a self-service experience from data prep to data delivery with minimal IT intervention, while putting in the necessary 'guardrails' for enterprise oversight and control.

CDAP BENEFITS

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

SOLUTIONS

- EDW Offload and CDC
- Managed Data Lake
- 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
- Ericsson
- Thomson Reuters
- 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.

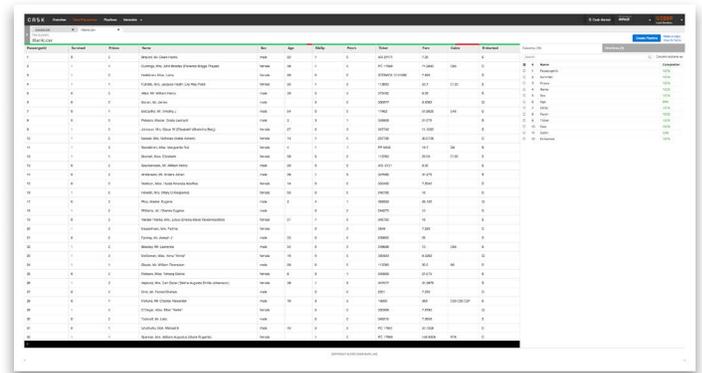
Cask Market

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.



Visualize, Transform, and Wrangle Data

CDAP provides an easy and interactive way to cleanse, transform, and blend your complex datasets. It makes data transformation not only easier, but faster and more fun, too. CDAP 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.

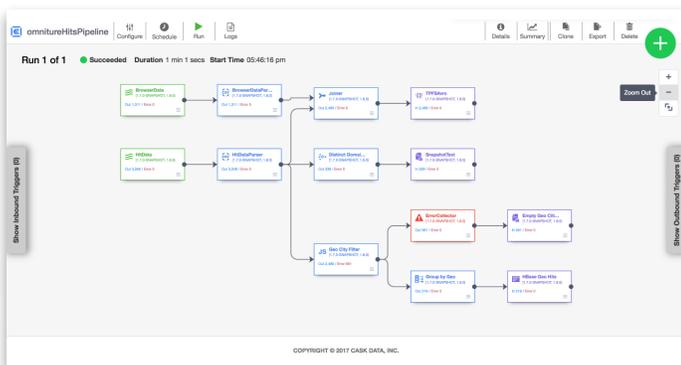


Putting Big Data at your Fingertips

CDAP is designed to solve common and critical big data challenges. It provides a visual interface with rich, pre-built libraries eliminating the need for manual coding, reducing overall complexity. Users can work seamlessly on all aspects of data integration and governance from data pipelines, to data wrangling, and metadata management. Additionally, the security, monitoring, and management of Hadoop and Spark installations is vastly simplified to enable an enterprise-ready production environment.

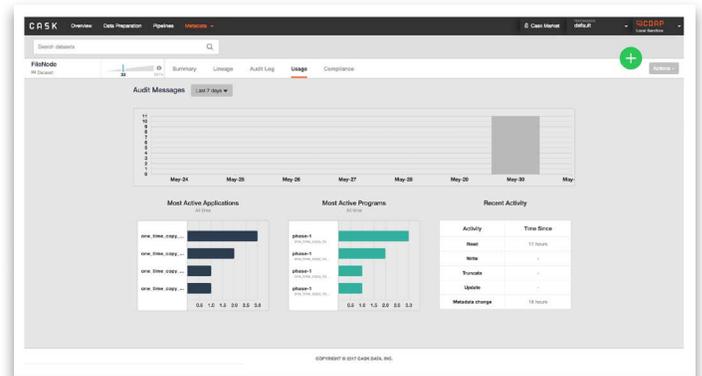
Build & Manage Complex Data Pipelines

CDAP provides a data ingestion framework that simplifies and automates the difficult and time consuming task of building and managing data pipelines. It 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. Spark data pipelines can be built and run interactively.



Automated Capture of Metadata & Lineage

CDAP helps you discover, profile, and govern data entering 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. CDAP allows you to perform impact and root-cause analyses, delivering an automated and comprehensive audit trail for compliance.



Summary

CDAP, with its built-in tools and Cask Market big data app store, provides businesses with a single, unified integration platform that simplifies delivery of a new generation of data lakes and machine learning apps on Hadoop and Spark. Leveraging CDAP platform capabilities, Cask also offers the Cask Distributed Rules Engine and the Cask Microservices Framework to support additional use cases. Find out more about CDAP at cask.co.

