



CASK DATA APPLICATION PLATFORM

Integrated Framework for Building and Running Data Applications on Hadoop

Supports All Major
Hadoop DistributionsIntegrates the Latest
Big Data TechnologiesFully Open Source
and Highly Extensible

Data Integration

- **Data Ingestion** Both realtime and batch, push and pull. Support for Flume, Kafka and REST.
- **ETL and Workflow** UI-based pluggable ETL framework with simple programmatic workflow and scheduling engine.
- **Data as a Service** Expose data via REST APIs for simple microservices or query with SQL/JDBC.



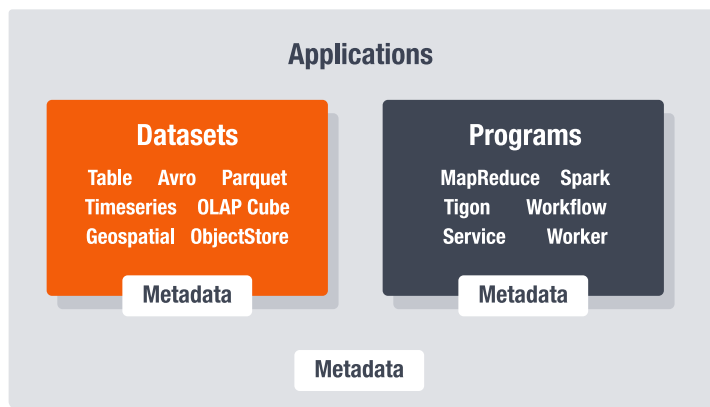
Application Development

- **Simple Java APIs** Provide application packaging, reusable data patterns and standardized application templates.
- **Local SDK** Standalone mode enables local development with IDE integration, rich UI and interactive shell.
- **Realtime and Batch** Transactions provide framework-level correctness and guaranteed consistency.



Production Operations

- **Standardized Model** Datasets, Programs, Logs and Metrics.
- **Runtime Services** Deployment, Discovery and Monitoring. Metadata and Security.
- **DevOps Tools** Rich UI Console, Interactive Shell, full REST APIs.



• Application Container Architecture

Standardized packaging of datasets and programs into application artifacts to simplify the application lifecycle and enable easy adoption of new technologies.

• Reusable Programming Abstractions

Powerful abstractions to encapsulate domain-specific data and application patterns that can be consumed by application developers and reused across applications.

• User and Machine Metadata

Rich metadata for every object combines arbitrary user metadata with automatically generated machine metadata based on all interactions of objects.

• Unified, Consistent Access to Data

Consistent access to all data in realtime, batch and SQL with transactional guarantees.

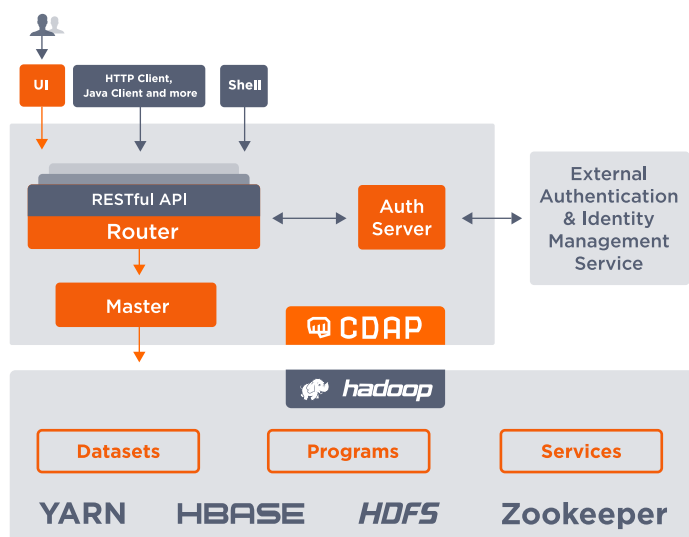
• Application and Data Portability

Application portability enables movement of apps and data between environments (dev, test, CI, prod) and distributions without any code changes.

• Production Runtime Environment

Production environment provides high-availability and horizontal scalability with support to manage, monitor, administer and secure applications and data.

Test	Local	Distributed
Threads NavigableMap Local Temp Files	Threads LevelDB Local Files	YARN HBase HDFS
		amazon Google Microsoft



CDAP Server

- Highly Available (HA)
- Installed on edge node(s)
- Supports Kerberos impersonation and perimeter security
- Manages system services in YARN
- Provides access point for all cluster interactions

System Services Apache Twill Containers

Powered By Apache Twill

- Transactions (Tephra)
- Metrics Aggregation
- Log Aggregation
- Dataset Services
- Metadata Management
- Explore Service
- Stream Management and more

Try CDAP Today



Download

Download the CDAP SDK or VM to try CDAP on your laptop

<http://cask.co/download>



Get Started

Try one of the many sample applications, tutorials or user guides

<http://cask.co/get-started>



Learn More

Cask provides commercial support and services for CDAP

sales@cask.co