CDAP
CASK DATA APPLICATION PLATFORM

Quickly and easily develop and deploy more powerful applications on Hadoop

The Cask Data Application Platform (CDAP) is an open source integrated platform for developers and organizations to build, deploy, and manage data applications

Value of the Cask Data Application Platform

Drive greater innovation with data
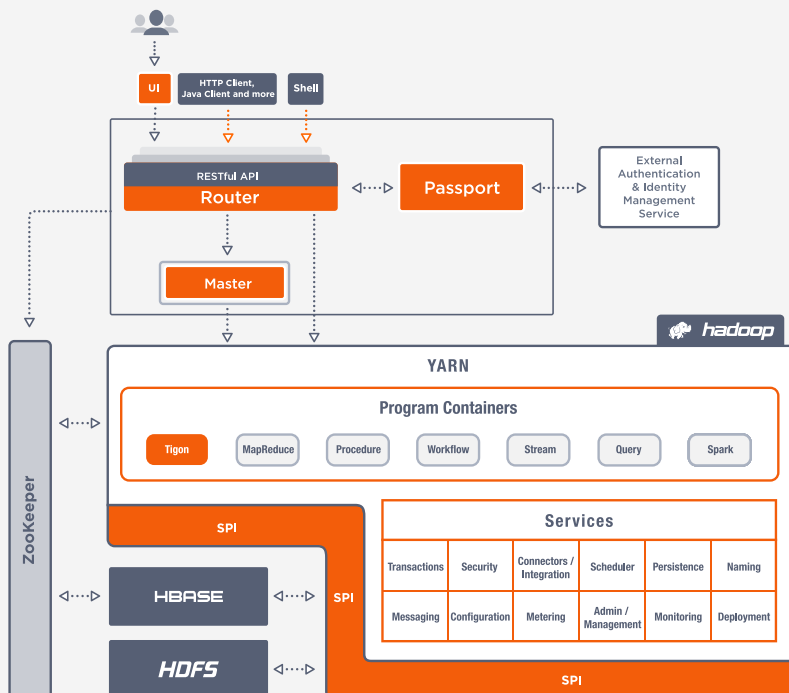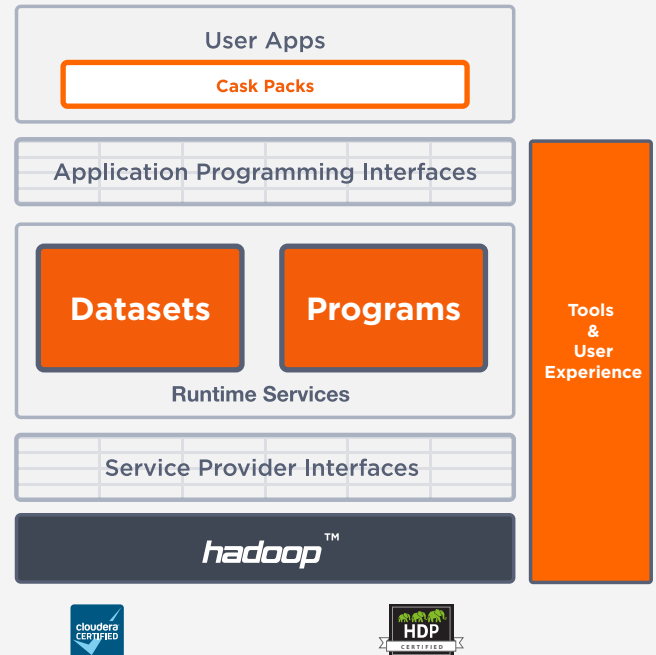- New class of applications to increase business value, including those requiring real-time and batch processing

Accelerate and simplify development
- Utilizing existing talent and development processes

Manage your application life cycle
- Enterprise best practices for development, debugging, deployment, and maintenance

Functional architecture

User Apps
Cask Packs
Application Programming Interfaces
Datasets
Programs
Runtime Services
Tools & User Experience
Service Provider Interfaces
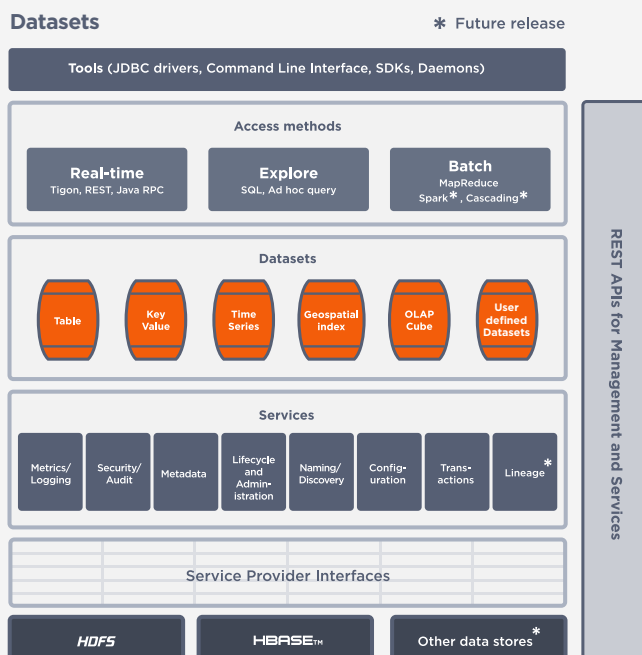hadoop™

System capabilities

Abstraction
- Support multiple programming paradigms and data stores interchangeably

Integration
- Expose low level functions as higher level capabilities

Runtime services
- Includes services for apps and data such as security, discovery, and management throughout the app and data lifecycle

System architecture

UI · HTTP Client, Java Client and more · Shell · RESTful API · Router · Passport · External Authentication & Identity Management Service · Master · Zookeeper · hadoop · YARN · Program Containers · Tigon · MapReduce · Procedure · Workflow · Stream · Query · Spark · SPI · HBASE · HDFS · Services · Transactions · Security · Connectors / Integration · Scheduler · Persistence · Naming · Messaging · Configuration · Metering · Admin / Management · Monitoring · Deployment

Datasets

Tools (JDBC drivers, Command Line Interface, SDKs, Daemons)

Access methods

Real-time	Explore	Batch
Tigon, REST, Java RPC	SQL, Ad hoc query	MapReduce Spark*, Cascading*

Datasets

Table | Key Value | Time Series | Geospatial index | OLAP Cube | User defined Datasets

Services

Metrics/ Logging | Security/ Audit | Metadata | Lifecycle and Administration | Naming/ Discovery | Config-uration | Trans-actions | Lineage*

Service Provider Interfaces

HDFS | HBASE™ | Other data stores*

REST APIs for Management and Services

Libraries to build reusable data access patterns spanning multiple storage technologies

Programs

CDAP APIs | **Other Open Source APIs**

App Containers

Workflow | Tigon Flows | Procedures | User Defined Services | Spark* | Map Reduce

Services

Metrics/ Logging | Security/ Audit | Metadata | Lifecycle and administration | Naming/ Discovery | Config-uration | Trans-actions | Datasets

Service Provider Interfaces

hadoop™

REST APIs for Management and Services

Standardized containers providing consistency for diverse processing paradigms

Features

Streams for data ingestion

- Supports Kafka, Flume, REST, user-defined protocols
- Time-stamped and ordered
- Horizontally scalable

Logical representations in commonly used access patterns

- Time series, Key value, objects, geospatial index, OLAP cube and more
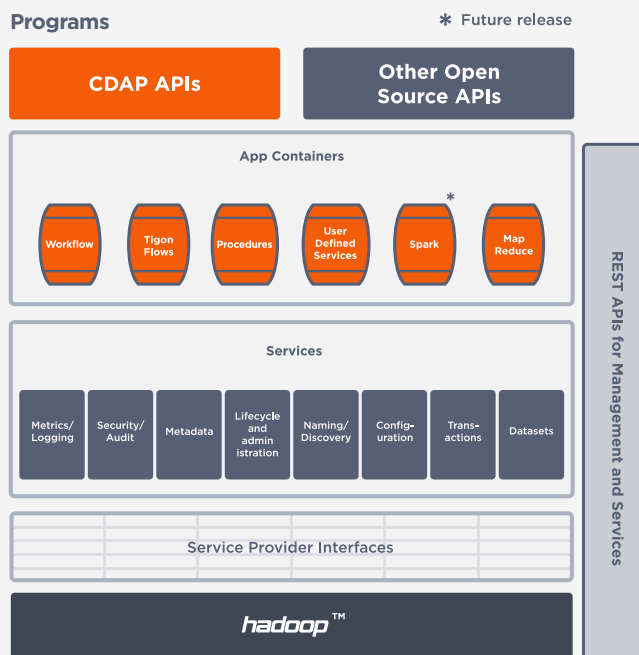
Data available to multiple applications

- MapReduce, Hive, Spark, Flows and more
- REST APIs
- Unified batch and real-time processing

Features

Framework level guarantees

- Applications aren't required to be idempotent

Support for development life cycle and operational deployment

- Portable from laptop to cluster
- Logging, metrics, security with no developer overhead

Standardization of containers across programming paradigms

- Take advantage of Spark, Cascading, Hive, etc. using programmatic APIs without worrying about system implementation

Benefits

- Simplify data ingestion and Extract Transform Load (ETL) to accelerate time to value
- Maximize value of data by making it easy to find and easy to explore through multiple query methods
- Protect data through security, audit, lineage, and reporting

Benefits

- Developers can build a broader range of apps focusing on business logic, not core system services
- Speed up time from coding to testing to operational deployment
- Take advantage of new technology with less need for training and expertise

Product options

	Community Edition	Development Subscription	Production subscription
Scope of use		Application development	Application and production
Pricing	Free	Annual subscription based on CDAP server count	
Support	Community Support	24x7 Support portal 8x5 live support 1 day response	24x7 Support portal 24x7 live support 4 hour response
Indemnity	⊖	✓	✓
Updates	Notification	Updates installed by Cask	

Download options

CDAP Virtual Machine

Get running quickly with CDAP Standalone in a fully pre-configured Linux VM environment.
Requires VirtualBox. Windows, Mac, Linux

CDAP Standalone

Develop and deploy Hadoop data and apps with a fully functional CDAP environment designed to run on your laptop.
Requires Java and NodeJS. Mac, Linux

CDAP Distributed

Run the fully distributed, highly available YARN, HDFS, and HBase based version of CDAP for full scale testing and production.
Requires Hadoop 2.0+. Linux

Download from http://cask.co/resources/#cdap