Hadoop Enables New Applications and Architectures

ENTERPRISE DATA LAKES

Batch and Realtime Data Ingestion

Any type of data from any type of source in any volume

Batch and Streaming ETL

Code-free self-service creation and management of pipelines

SQL Exploration and Data Science

All data is automatically accessible via SQL and client SDKs

Data as a Service

Easily expose generic or custom REST APIs on any data

BIG DATA ANALYTICS

360° Customer View

Integrate data from any source and expose through queries and APIs

Realtime Dashboards

Perform realtime OLAP aggregations and serve them through REST APIs

Time Series Analysis

Store, process and serve massive volumes of time-series data

Realtime Log Analytics

Ingestion and processing of high-throughput streaming log events

PRODUCTION DATA APPS

Recommendation Engines

Build models in batch using historical data and serve them in realtime

Anomaly Detection Systems

Process streaming events and predictably compare them in realtime to historical data

NRT Event Monitoring

Reliably monitor large streams of data and perform defined actions within a specified time

Internet of Things

Ingestion, storage and processing of events that is highly-available, scalable and consistent
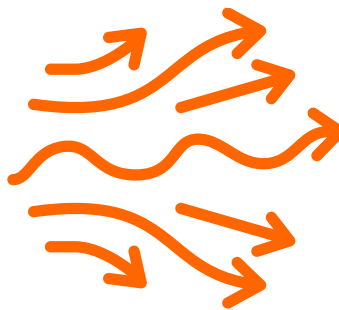
Data Applications Drive Meaningful Business Value

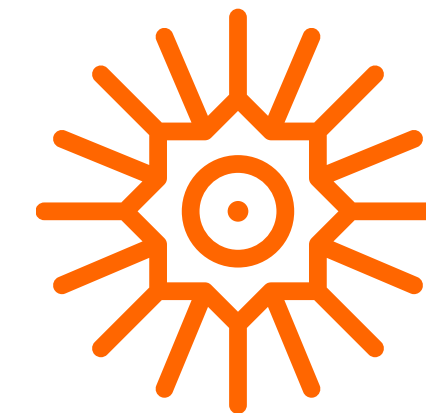# But Getting Value from Big Data is Hard
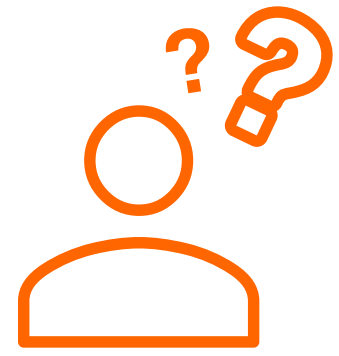
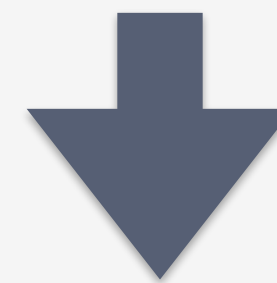**Complexity** of technologies and new user learning curve

Proliferation of projects, services and APIs

Divergence of distributions and technologies

Integration silos created by narrow point solutions

Too much focus on infrastructure and integration, rather than applications and analytics

And There Are Many Faces of Hadoop

Developer

Architecture & Programming

Focused on Apps & Solutions

Ops

Configuring & Monitoring

Focused on Infrastructure & SLA's

Data Scientist

Scripting & Machine Learning

Focused on Data & Algorithms

LOB / Product

Driving Revenue & Decision Making

Focused on Products & Insights

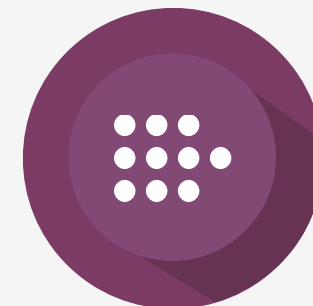**Without a consistent set of tools, IT will not be an effective data enabler for the business**

Enter Cask

Founded in 2011
By early Hadoop engineers from
Facebook and Yahoo!

Raised $37+ Million
Andreessen Horowitz, Safeguard,
Battery Venture and Ignition Partners

Strategic Investors
AT&T, Cloudera and Ericsson

Latest Release
3.6 Cask Data Application Platform,
Cask Hydrator and Cask Tracker

Key Customers & Partners
AT&T, Ericsson, Lotame, Salesforce, Cloudera,
Hortonworks, MapR, Microsoft, IBM, Tableau…

NEW: CDAP 4 Preview
Featuring Cask Market,
the "big data app store"

Why "Cask" ?
A Container Architecture that puts
Big Data on Tap

The Evolution of the Cask Platform

Convergence of Big Data Apps and Data Integration →

CDAP v2	CDAP v3	CDAP v4

Big Data App Server

- Abstractions & integrations
- Metrics & logs
- Debugging environment

Big Data Apps + Data Integration

- Data ingest
- Data pipelines
- Workflows and metadata

Unified Integration for Big Data

- Security & governance
- Self-service environment
- Enterprise integrations

"WebLogic for Hadoop"   **"WebLogic Meets Informatica"**   **"Unified Big Data Integration"**

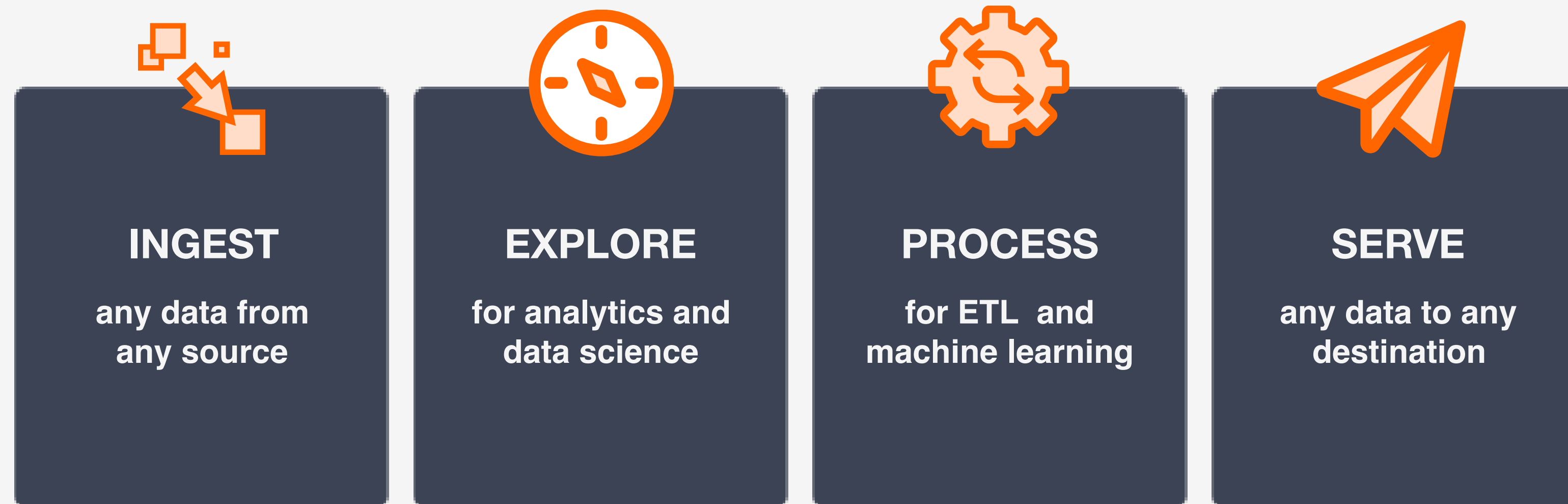# Introducing Cask Data Application Platform (CDAP)

First Unified Integration Platform for Big Data

Platform for distributed apps, bringing together application management with data integration

- 100% open source and built for extensibility

- Supports all major Hadoop distributions and clouds

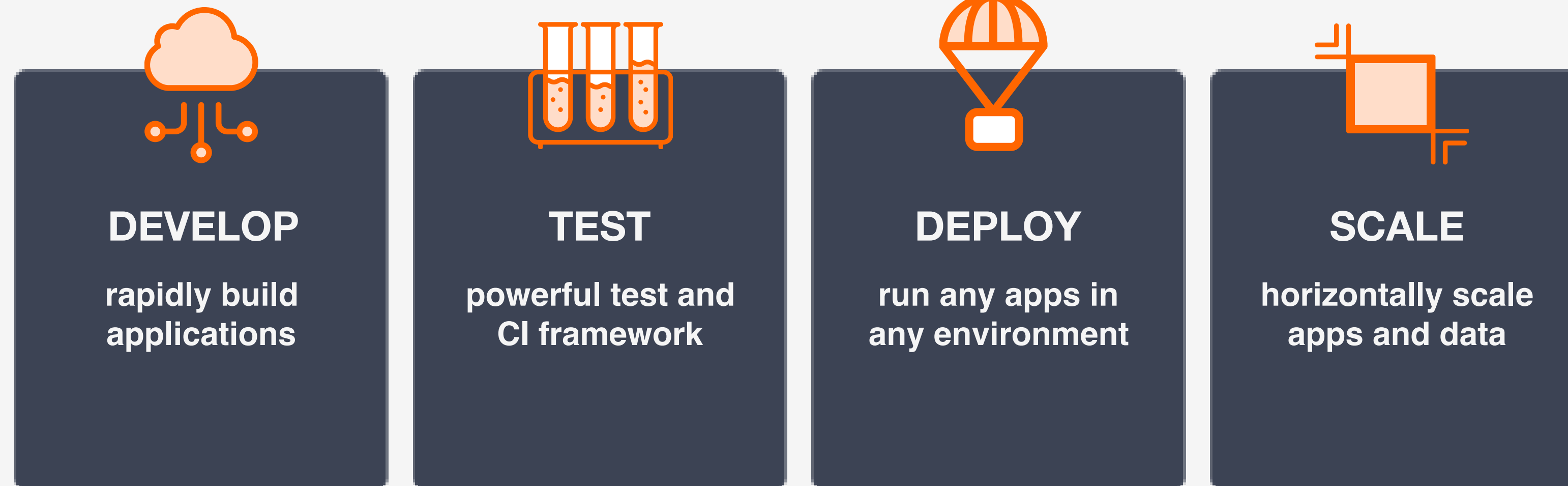- Integrates the latest open source big data technologies

⦿CDAP

Security and Governance

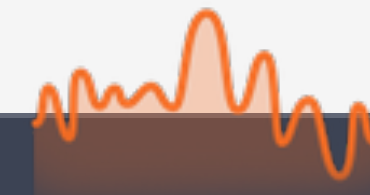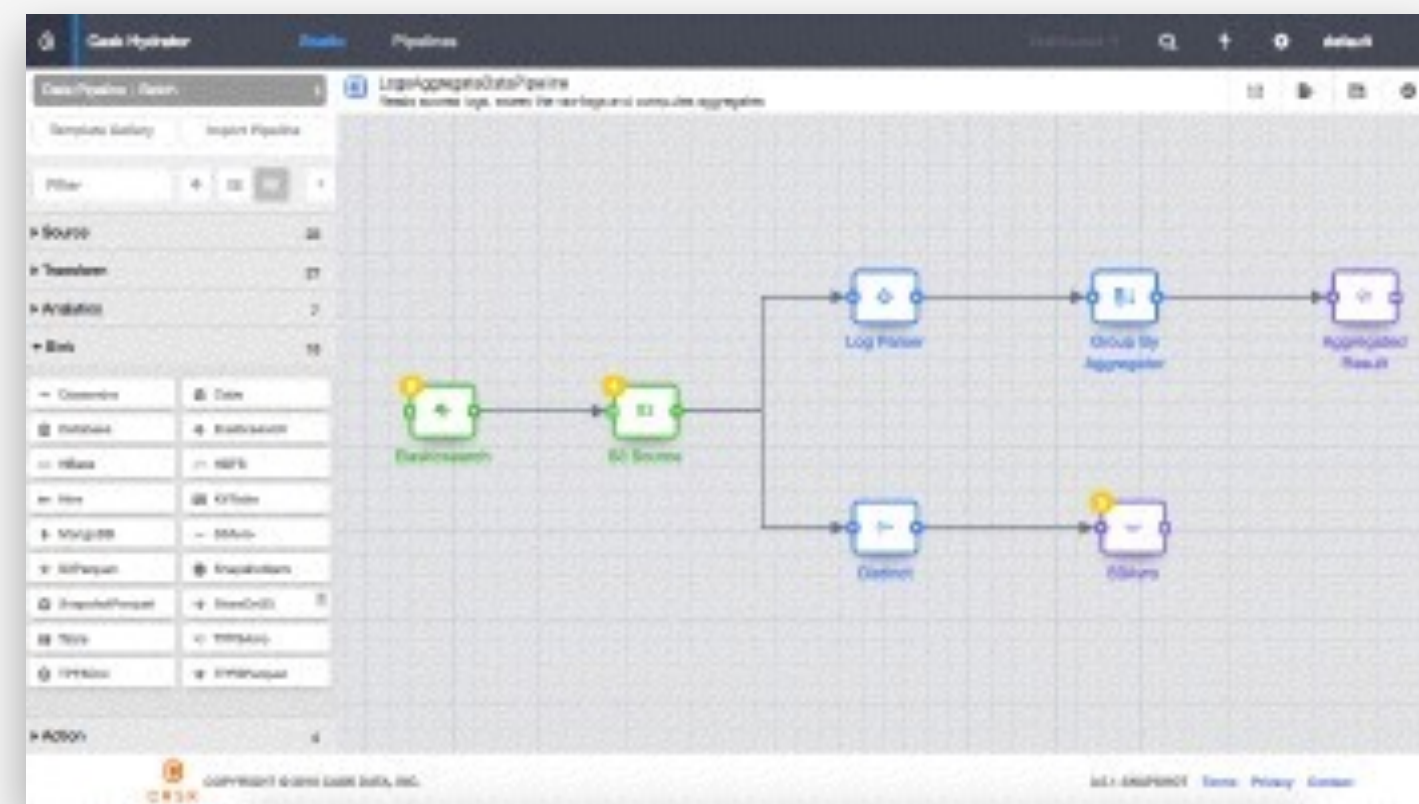| CAPTURE	DISCOVER	TRACK	ANALYZE
store all metadata about your data	easily locate any of your data	every audit plus lineage graphs	understand usage patterns of data

ENCRYPT   AUTHENTICATE   AUTHORIZE

CASK

CDAP

Self-Service User Experience

Cask Hydrator

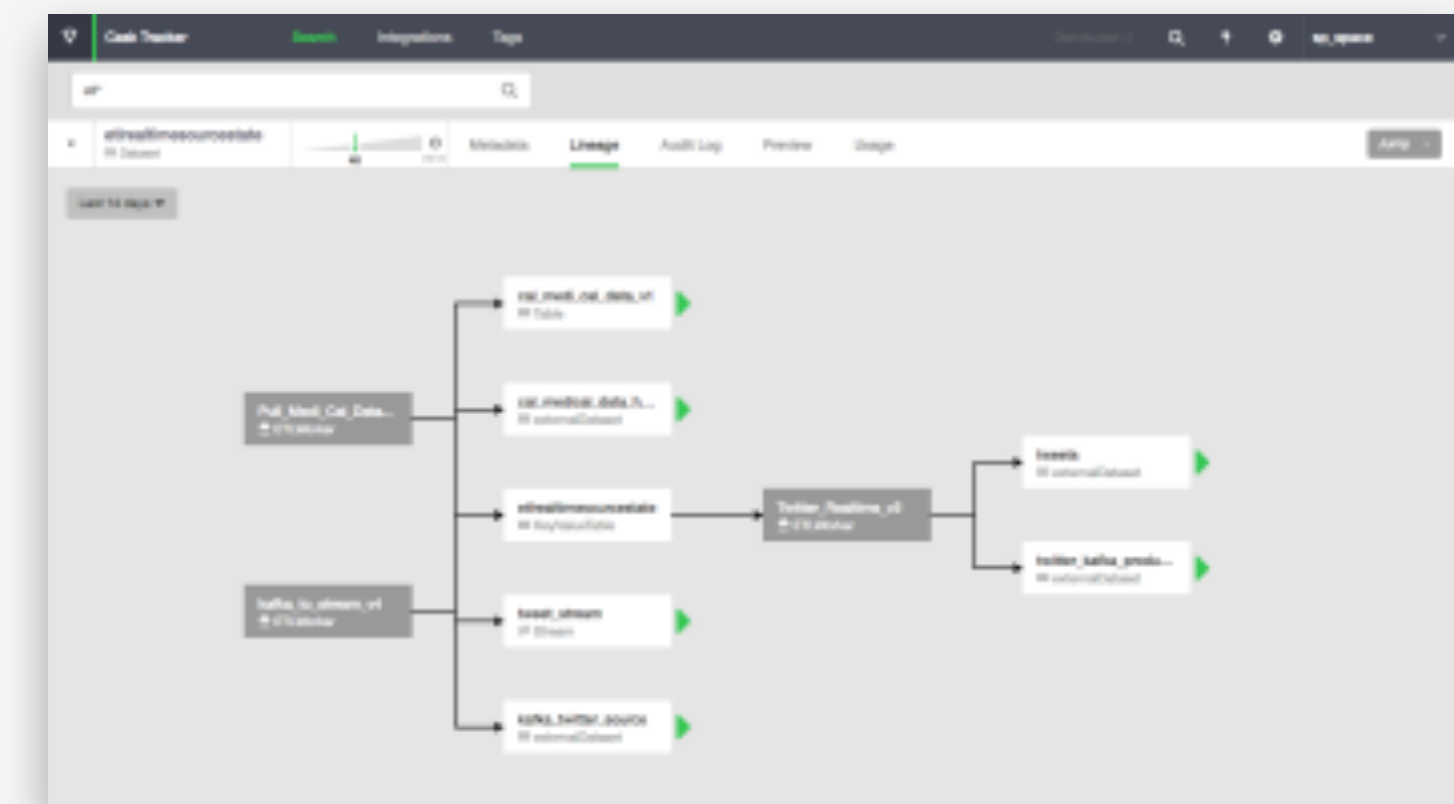**A code-free framework to build and run data pipelines**

| Drag & drop graphical interface | Create, debug, deploy and manage | Separation of logic and execution environment | Native to Hadoop & Spark — scales out |

Cask Tracker

A data discovery tool to explore metadata and usage

| Rich app-level metadata | Track lineage and audits | Analyze usage of datasets | MDM integration framework |

cask.co

cask

The CDAP Architecture

Applications

Datasets

Table   Avro   Parquet

Timeseries   OLAP Cube

Geospatial   ObjectStore

Metadata

Programs

MapReduce   Spark

Tigon   Workflow

Service   Worker
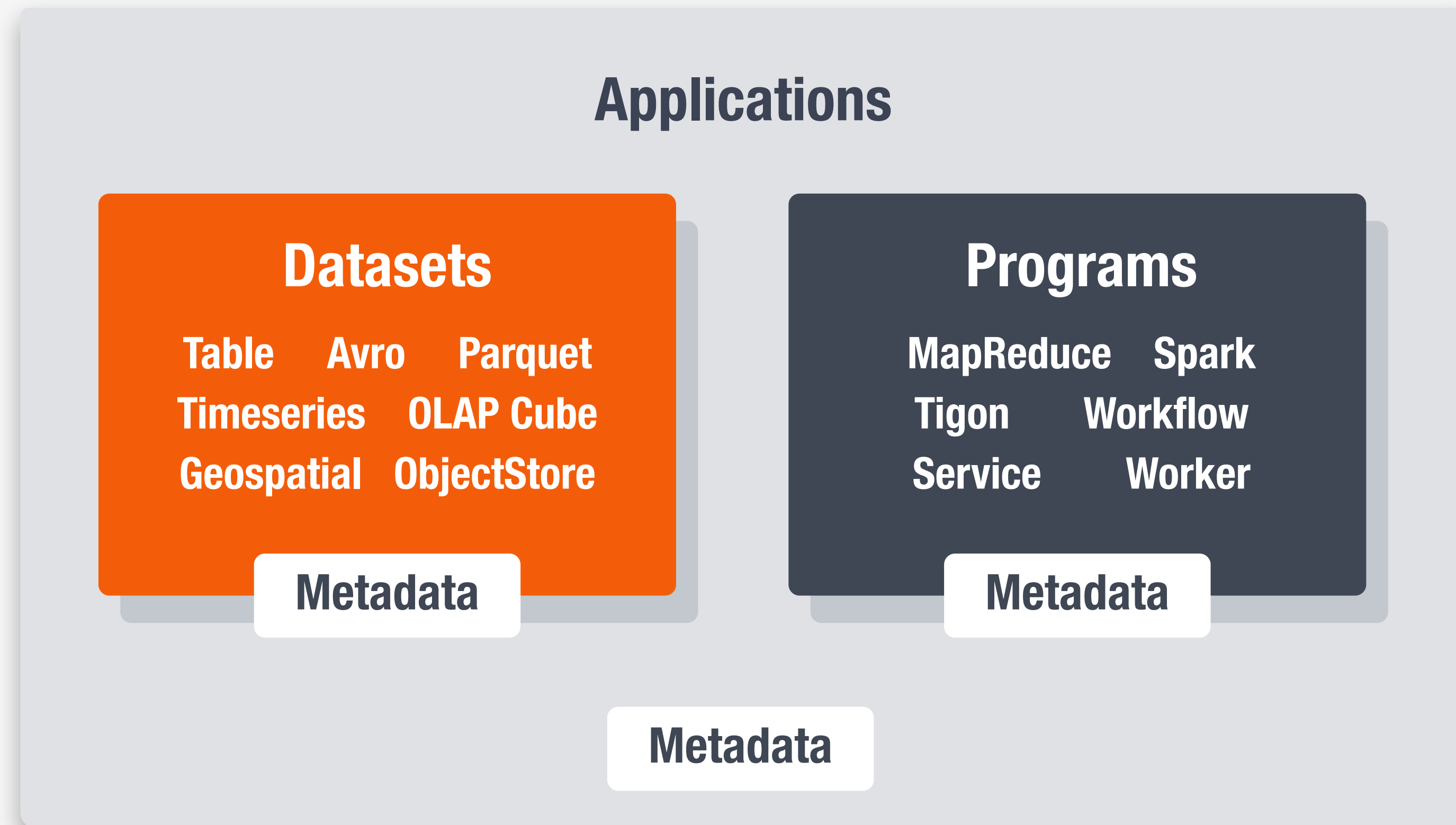
Metadata

Metadata

- **Application Container Architecture**
- **Reusable Programming Abstractions**
- **Global User and Machine Metadata**
- **Highly Extensible Plugin Architecture**

CASK

CDAP Enables the Full Big Data Application Lifecycle

- **Standardization,** deep integrations, tools and docs
- **Separation** of app logic from data logic and integration logic
- **Conceptual integrity** within applications and consistency across environments

Rapid Development

Production Operations & Governance

- **Simplified** packaging, deployment and monitoring of apps on Hadoop
- Enhanced security and governance with **centralized metrics and logs**
- **Tracking and exploration** of metadata, data provenance, audit trails and usage analytics

Single framework for building and running data apps and data lakes on Hadoop and Spark

⬇

CDAP

reduces time to develop and deploy big data apps by 80%

reduces time to insights and accelerates business value

removes barriers to innovation and future-proofs your apps

Customer Success Stories

Customer Situation

Health Insurance Provider
offloading clinical / immunization reporting from Netezza

Lack of existing Hadoop expertise and frustration with hand-coding and scripting tools

Leading SaaS Platform
taking new real-time, massive scale products to market

Small team and significant technical challenges limit pace of development and solution scale

Large Telco Enterprise
building a centralized, secured, multi-tenant Data Lake

Multiple teams and technologies with widely varied skillsets and incompatible design choices

---

Cask Solution

Cask Hydrator for rapid creation of data pipelines and Cask Tracker for data discovery

POC in 2 days
Production in 2 months

CDAP for real-time ingestion and consistent processing with production operations support

Development in 1 month
Production in 3 months

CDAP for data lake management and orchestration, tightly integrated into existing systems

Hundreds of Users
Thousands of Pipelines

Awards and Accolades

Cask was Named a Gartner Cool Vendor 2016

Cask was Certified a Great Place to Work 2016

" … "**CDAP is a big win for us** … the amount of code we needed to write was minimal with CDAP, and it was much easier and faster than we ever expected …"

(Jia-Long Wu, Data Architect, Lotame)

" … "**Cask has tilted the playing field**, earning a massive unfair advantage over proprietary point products for data integration and ingest …"

(Nik Rouda, Senior Analyst, Enterprise Strategy Group)

" … for the rest of us who lack the technological chips or patience to make it all work, there's good news: it will soon get easier, **thanks to** the work done by the big data pioneers, as well as **vendors like Cask** …"

(Alex Woodie, Managing Editor, Datanami)

NEW: CDAP 4 — Big Data Apps on Tap!

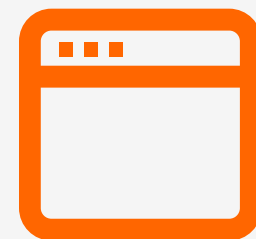## Release of CDAP 4 Preview

Available for download now!

Cask Market
"Big Data App Store"

Cask Wrangler
Interactive Data Preparation

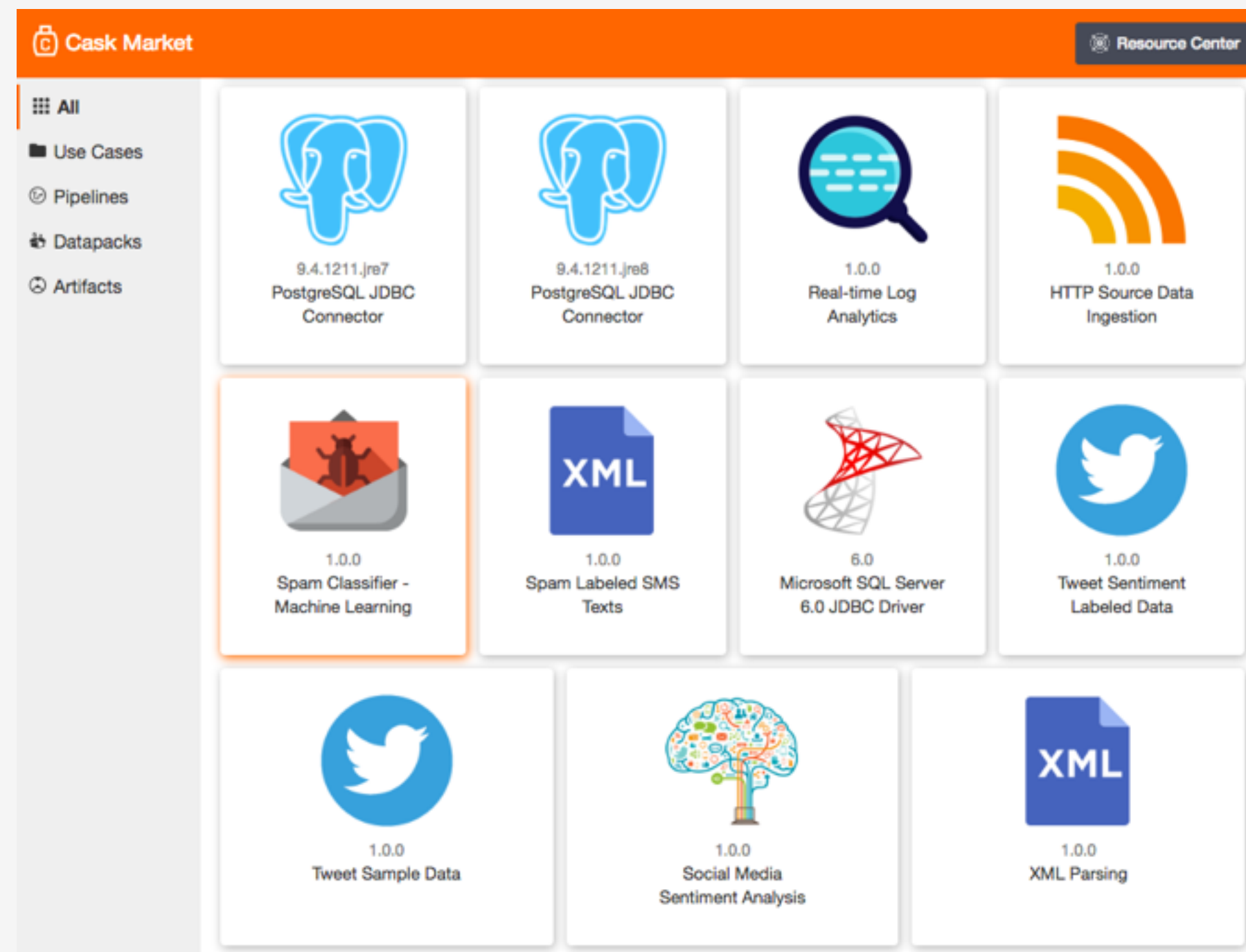### Reimagined CDAP UI
Rewrite based on React

Resource Center
Interactive Wizards for Common Tasks

NEW: CDAP 4 — Big Data Apps on Tap!

Cask Market

The "App Store for Big Data"

- **Goal: Time to value in minutes w/ <u>no existing experience</u>**
- Application and Library Ecosystem with pre-built Hadoop solutions, reusable templates, and third-party plugins
- Available from anywhere inside the CDAP UI with a click

- Initially, everything in the Cask Market has been bootstrapped by Cask based on ongoing work across our customers, is 100% open source and available on GitHub
- Eventually, developers and ISVs will be able to showcase and market their own applications and libraries (ex: Graylog)

Cask Market includes Interactive, Guided Wizards for Configuring Pre-Built Templates

C A S K

Summary

- CDAP provides the **first unified integration platform for big data**

- Cask Hydrator and Cask Tracker are visual extensions of CDAP for **self-service access**

- CDAP empowers enterprise IT to deliver **faster time to value** for Hadoop and Spark, from prototype to production

- **Cask Market** is a "big data app store" available in CDAP 4 with pre-built apps, pipelines, plugins

- CDAP is **100% open source**, highly extensible, enterprise-ready, and commercially supported

C A S K

Big Data on Tap

cask.co

Thanks!

For more information, go to: **cask.co**

CASK