



Key Facts

Apache Committer Seats Across Ecosystem

118 (as of February 2016)

PMC Seats Across Ecosystem

81 (as of February 2016)

Projects Founded by Cloudera Employees

Apache Avro
 Apache Bigtop
 Apache Crunch
 Apache Flume
 Apache Hadoop (co-founder)
 Apache HBase
 Apache Hive (co-founder)
 Apache Impala (incubating)
 Apache Kudu (incubating)
 Apache Lucene/Solr
 Apache Sentry (incubating)
 Apache Sqoop
 Apache Parquet (co-founder)
 Apache ZooKeeper
 Hue
 Ibis
 Oryx
 Pandas
 RecordService

Cloudera's Approach to Open Source and Open Standards

When Cloudera's chief architect Doug Cutting founded the Apache Hadoop project, it was with an open source vision firmly in mind. Throughout its history, Cloudera has been strongly committed to a community-driven, Hadoop-based platform based on open standards that meets the highest enterprise expectations for stability and reliability.

Today, Cloudera's approach to open source and open standards has one goal: customer success. And that goal is achieved by:

Innovating in open source

Some vendors *consume* the open source community's activity, and others help *drive* it. Cloudera shapes the evolution and roadmap of the Hadoop platform by creating, contributing, and supporting new enterprise-focused capabilities.

- Cloudera has contributed, and shipped, more major features to the platform than any other company.
- Contributions include HDFS NameNode HA, network & HDFS encryption, fine-grained access control, and Spark-YARN integration.

Curation of open standards

Open standards are components that attract widespread ecosystem investment due to universal adoption—and thus are the best foundation for mainstream, long-term architecture.

- Cloudera has a proven track record of identifying (and often creating), curating, and supporting standards, including Apache HBase, Apache Impala (incubating), Apache Spark, and Apache Kafka.
- Every component inside CDH is a standard (that is, is shipped by multiple vendors), so your investments are safe.

Highest enterprise requirements

Cloudera leads the way to support the *entire* open source platform for customers, not just the core, and maintains a persistent focus on platform stability and reliability.

- Cloudera employees collectively hold more than 100 Apache committer seats across the platform, ensuring timely and permanent fixes, not just workarounds.
- Cloudera invests significant and dedicated engineering resources in multi-dimensional testing and additional quality assurance before any release is done.

The CDH Life Cycle

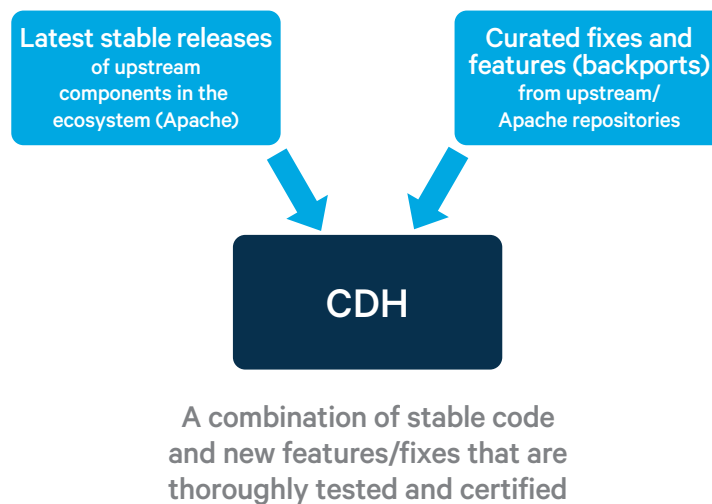
Major Releases

Each major release of CDH (aka CDH X) begins with inclusion of the latest stable releases of Apache components after extensive testing, integration, and tuning (fit-and-finish). In cases where functionality is not production-ready or compatibility is broken across those major releases, we'll skip the problematic parts—choosing instead to curate critical bug fixes and features and backport them into CDH. (For example, due to backward incompatibility across Apache Hive 0.10 and 0.11, Cloudera never shipped the latter in its entirety.)

Minor and Point Releases

Thanks to the broadest customer and partner feedback channels in the industry, Cloudera's Apache committers are also continually writing new bug fixes and contributing them to the project trunks upstream. (Cloudera has an "upstream-first" policy; patches always go there as a first step.)

Users who rely on Apache exclusively have to wait for an official Apache release to get access to those patches—and when they do, their only option is to consume the entire patchset, regardless of their impact on existing applications. In contrast, for CDH users, every four months critical patches are selectively aggregated and backported to CDH and made available in the form of minor releases (aka CDH X.Y)—with some very critical ones shipping as "dot-dot" releases, as well. In all cases, Cloudera is diligent about ensuring that these patches don't alter application behavior.



Users get the best of both worlds: stable, released code in combination with curated, forward-looking features and bug fixes. The advantages being:

- Users can confidently access new Apache releases that have been validated via extensive unit and integration testing.
- User can count on their issues being fixed permanently upstream.
- Users can consume the most critical new upstream bug fixes and innovations at a regular cadence.
- Interoperability and stability is ensured across releases, as well as with the upstream project trunks.
- Upgrades are significantly easier.

This approach has been validated time and time again by Cloudera's users as the best option for enterprise-class deployments. And if they're successful, so are we.