

HADOOP RESOURCES: TRAINING, CONFERENCES & MORE



Where to learn more and stay informed

This brief online guide to Hadoop was intended to be informative and accurate, but not comprehensive. It doesn't claim to teach you everything you need to know about Hadoop, but does highlight what you need to know and where you can go to learn more. There are many excellent resources for learning more about Hadoop technology and how to use it. Some of the most credible and informative sources are highlighted in this section.

Must-read Hadoop books

There are several excellent books on big data in general and Hadoop in particular.

For an in-depth technical read on Hadoop, try ***Hadoop: The Definitive Guide*** by Tom White. It serves a programmer audience and has individual chapters covering MapReduce, HDFS, YARN, Spark, Flume, Crunch and other Hadoop components and related Apache projects.



Big Data: A Revolution That Will Transform How We Live, Work, and Think by Viktor Mayer-Schönberger and Kenneth Cukier, originally published in 2014. Written for a business/executive audience, it provides an overview of what big data is, with lots of use cases to help put big data into context. This book is more about big data, not focused on Hadoop specifically.

Two other good books also focused on the philosophy, use cases and value of big data:

- ***The Signal and the Noise: Why So Many Predictions Fail – But Some Don't*** by Nate Silver, who is known to many for his blogging and other writing on sports and politics, delves into using data for predictive forecasting.
- ***Data Smart: Using Data Science to Transform Information into Insight*** from data scientist and author John W. Foreman presents readers with the fundamentals of data science, including clustering, datasets, predictive analysis, forecasting and process optimization. It includes Excel tutorials to familiarize users with working with small datasets – without requiring coding knowledge.

[O'Reilly](#) also offers a number of books and guides, business and technology oriented, as well as training videos that support and educate specific to big data and Hadoop.

Hadoop conferences and events

Industry events provide the opportunity to attend conference sessions and to meet with industry experts and vendors to ask specific questions and learn about the latest technology developments. The content and size of events varies from year to year, but those listed below have consistently provided a quality experience for attendees looking to learn more about big data and Hadoop. Note that while the face-to-face portions of these events last only a few days, some have websites and blogs that are regularly updated with new resources.

[Strata + Hadoop](#) World brings all the major Hadoop distribution vendors and numerous other related software and services providers together under one roof and regularly draws thousands of attendees. It also includes an extensive educational track with conference sessions and in-depth technical training classes plus various other forums, keynote addresses, demos and other special events.

Hadoop Summit events are produced by Hortonworks and are held several times a year at various global locations. The June event in San Jose is the flagship. Hadoop Summits include a mix of technical- and strategic-focused conference sessions, including sessions that are selected based on voting by the community, HDP training and certification courses, an exhibit hall that features many Hadoop ecosystem solution providers, community meet-ups, and other networking and social events.

The IEEE Big Data Congress is oriented to engineers involved in services computing. It is a technical conference where papers and in-depth sessions are presented, and doesn't have the large trade show component of Strata + Hadoop World or the Hadoop Summit. The IEEE Big Data Congress is co-located with a series of other IEEE events dedicated to cloud computing, web services, mobile services and related technologies.

Certification courses – as noted in the Careers section, several Hadoop vendors offer their own certification and training courses in various formats.

Hadoop blogs and websites

Several excellent blogs regularly cover Hadoop and big data. Blogs don't typically provide the depth of information available through conferences and books, but have the advantage of being timely. They are a good way to learn about new developments. Don't overlook the comments sections, which are often an excellent source of different points of view and troubleshooting tips.

[Dataconomy](#) is a website focused on data science that also has a newsletter and various companion social media sites and streams. Many expert contributors produce content which is usually not highly technical and focuses on the business and strategy aspects of big data.

[O'Reilly Radar](#) is a blog managed by the organization that runs Strata + Hadoop World and other events. Far from a promotional engine for the shows, the blog provides a lot of practical Hadoop technical advice and thought-provoking posts and podcasts about how to take advantage of big data.

[Gigaom](#) presents original research and other news about emerging technologies. It is not focused exclusively on big data and provides regular coverage of SMAC (social, mobile, analytics, cloud), [security](#) and Internet of Things (IoT) technologies.

[Silicon Angle](#) covers the business side of big data, with extensive reporting about vendor news, personnel moves, partnership agreements, funding announcements, new product releases and other developments. The site puts news into perspective with original reporting and analysis by knowledgeable staff writers.

Tech vendor sites and resources

As long as you're aware of the potential vendor bias, the commercial Hadoop distribution vendors and other ecosystem companies can be excellent sources of information. They provide blogs and communities that present new thinking, how-to technical information, troubleshooting advice, tutorials, videos, podcasts and other resources, plus links to partner and other sites. [The Cloudera community site](#), [Hortonworks blog](#), and [MapR blog](#) are good places to start.

Key Hadoop analysts

The analyst community follows Hadoop closely often focused on big-picture issues such as strategies for implementing Hadoop, best uses cases, new developments to be aware of, and what role Hadoop should play in your future IT and business structure. While most of the reports and in-depth analysis are only available through paid subscriptions, the firms give away a good amount of insightful and entertaining information through their blogs, newsletters, press releases and other outlets. Blog sections of the analyst websites are a good place to start browsing because they present recent insight and don't require you to sort through a long list of search results that includes both accessible and locked content.

Forrester Consulting has more than 100 [Hadoop-related blog posts](#) on its website, plus you can search the main site for terms like "Big Data" and "BI" to find additional, free perspectives.

IDC's [Big Data hub](#) puts press releases, tweets and links to its other social media channels plus announcements of new reports and other paid content in one place. IDC's various blogs are hosted at the [IDC Community](#) site, where you can search for Hadoop content. IDC press releases are searchable by its Big Data filter.

Ovum frequently reports on Hadoop and other big data topics. The easiest way to find relevant content is to do a search from the home page.

Hadoop thought leaders to follow

Doug Cutting and Mike Cafarella are credited with creating Hadoop. They are among the important thought leaders you can follow on Twitter ([@cutting](#) and [@MikeCafarella](#)). Several Hadoop experts have attracted large followings because of their insights and perspective. The following four have all attracted more than 50,000 followers:

Kirk Borne ([@KirkDBorne](#)), an evangelist for data literacy.

Vincent Granville ([@analyticbridge](#)) shares news and insights about big data. He leads a community for big data practitioners, [Data Science Central](#) that includes a Hadoop channel, [Hadoop360.com](#).

Bernard Marr ([@BernardMarr](#)) is known for thought provoking, often contrarian views on big data and how to use it in business. He is a consultant, speaker and author.

Gregory Piatetsky ([@KDNugg](#)) is a scientist and entrepreneur who's tweets include business perspective, technical advice and lots of links to other interesting resources. >We hope this guide has been helpful in your organization's Hadoop efforts. Perhaps it will help lead you to success and becoming a Hadoop thought leader that others follow!