

IS HADOOP DEAD? HOW KUBERNETES AND CLOUD-NATIVE COULD DISPLACE HADOOP



When you Google Hadoop, you'll get a range of articles arguing whether the technology is completely kaput, a necessity, or somewhere in between.

While Hadoop for data processing is by no means dead, Google shows that [Hadoop hit its peak popularity as a search term in summer 2015](#) and it's been on a downward slide ever since. Sure, maybe the people who are using Hadoop are Googling it a lot less – they probably know what they're doing. But people who don't yet use Hadoop are likely bypassing it altogether for other more flexible options or seeking out alternatives to Hadoop's weaknesses.

What is Hadoop?

If you're considering whether the death of Hadoop, you likely already know what it is, but here's a brief primer. Apache Hadoop is an [open-source framework that stores data and can run apps on clusters of commodity hardware](#). It's known for its enormous processing power, allowing it to handle limitless concurrent tasks because of its distributed computing model. Plus, data is fail-safe because Hadoop automatically creates data backups, so you lose nothing in the case of a failure.

The real reasons companies love Hadoop, though, are its flexibility and scalability. Hadoop can gather vast types of data – structured and unstructured, from inputs like social media, clickstream, internal collections, etc. Then, sifting through the data sets, Hadoop determines which data are useful and which are futile, all without converting the data into a single format.

When data is determined futile for current needs, there's no need to remove the data set, because Hadoop can store the unprocessed data indefinitely. Store as much data as you need for as long as you'd like. Perhaps in five years you want to analyze data that wasn't previous useful, you can. And with minimal storage costs because of its commodity hardware and it's open-source nature, it's super cost effective.

Drawbacks of Hadoop

So, Hadoop came of age in the last decade as technology waded into the world of big data, offering ways not only to store, but also to quickly, roughly analyze data. Today, we have plenty more options that serve purposes that Hadoop can't touch. As with all technology, Hadoop has drawbacks – and these can be steep.

First up, while Hadoop is king for big data, it's actually [very inefficient for smaller data sets](#), so it's no silver bullet for all your data problems. This is because MapReduce, the data processing center of Hadoop, is file-intensive, making it good for simple information requests that are divided into independent units, but not so good on iterative and interactive analytics, which is the direction of big data. Though Hadoop can combine, process, and transform data, it doesn't easily provide the output you likely need – visuals and reporting that result in true business intelligence.

Secondly, "secure" it not a word that describes Hadoop. In fact, Hadoop's security settings are disabled by default, so an experienced data analyst would need to install security measures, making it less friendly for newer programmers.

As for experience: though Hadoop runs in Java, one of the leading programming languages around the world – it's often too complicated for newbies to handle. Not only does a Hadoop programmer need to know Java, he must know Hadoop enough to know when not to use it. Handle with care, because it's not great production. (Additionally, because Java is so widespread, its frameworks may be significantly more vulnerable.)

Finally, Hadoop isn't great at real-time analytics. It's not even good. (Kubernetes is.) Because Hadoop utilizes batch processing, response time is slow.

Hadoop replacements

So how do you address big data processing in a secure, flexible, real-time environment? Do other cloud services replace what Hadoop used to do? There aren't one or two single replacements for Hadoop. Instead, there are two major disrupters: software workarounds or fixes that improve Hadoop and cloud innovations.

More and more, programmers are finding workarounds or fixes to Hadoop's problems of security and medium-skill programming. For instance, new tools speed up MapReduce functionality: [Spark can be mounted on top of MapReduce](#) to process data up to 100 times faster. Entry-level programmers who are likelier to know SQL are using that language to navigate on top of Hadoop, making hiring easier because SQL is a common, easy-to-learn language. Still, there are no tools that offer comprehensive data standardization, data management and data governance.

Reaching the cloud is making it much easier to forego Hadoop altogether. The cloud is more viable for a range of data tasks we used to employ Hadoop for, like maintenance tasks of server scheduling and administrating tasks like file systems and file storage. With so many options for the

cloud, from serverless apps to FaaS and database-as-a-service, you can now effectively use the cloud as a database – it can be just or more efficient than Hadoop. Then, you can place a stack of custom, third-party apps upon the database. This offers benefits like flexibility, ease of programming, easier migration to mobile, and even improved security.

A popular option is Kubernetes, which [clusters containers across public, private, and hybrid clouds](#), eliminating many previously necessary deployment processes. The open-source container orchestration technology is picking up major traction as developers overwhelmingly embrace container technology, which particularly helpful in [DevOps](#) environments. Kubernetes is ideal for cloud-native apps that require speed, flexibility, and scalability.

[With the speed of Kubernetes, companies can take on near-real-time data analysis](#), something that poor Hadoop and MapReduce just can't offer. The goal of Kubernetes two-fold: to ingest huge amounts of data and understand the data in real-time, so companies can respond accordingly.

And that interest in real-time analytics is soaring. A comparison of Google search results indicates that [Kubernetes is on the rise just as sharply as Hadoop is on the decline](#).

So, is Hadoop dead? Like any technology, Hadoop won't solve all your problems, but it can be the right solution for the right environment. Still, cloud options may offer wider, deeper, and easier solutions right from the get go.