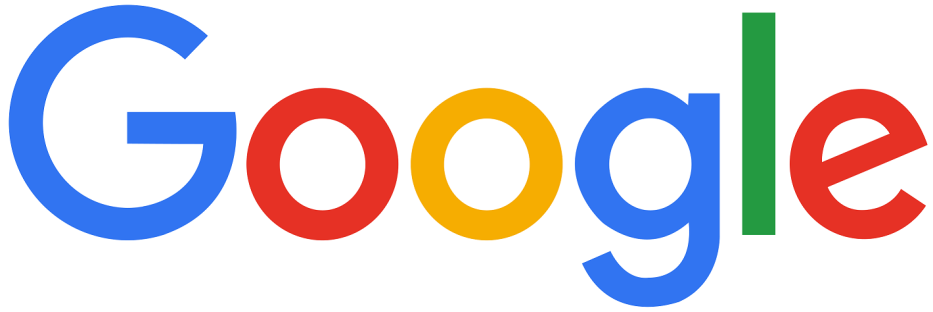
**Hadoop Ecosystem Explained**

**History:**

## ****How It All Started?****

So, it all started with two people, **Mike Cafarella** and **Doug Cutting,** who were in the process of building a search engine system that can index 1 billion pages. After their research, they estimated that such a system will cost around **half a million dollars** in hardware, with a monthly running cost of **$30,000,** which is quite expensive. However, they soon realized that their architecture will not be capable enough to work around with billions of pages on the web.



They came across a paper, published in **2003,** that described the architecture of [**Google’s distributed file system,**](https://static.googleusercontent.com/media/research.google.com/en/archive/gfs-sosp2003.pdf) called**GFS,** which was being used in production at [**Google**](https://www.google.org/)**.** Now, this paper on [**GFS**](https://static.googleusercontent.com/media/research.google.com/en/archive/gfs-sosp2003.pdf) .Later in 2004, Google published one more paper that introduced **[MapReduce](https://static.googleusercontent.com/media/research.google.com/en/archive/mapreduce-osdi04.pdf" \t "_blank)** to the world. Finally, these two papers led to the foundation of the framework called “**[Hadoop](https://hadoop.apache.org/docs/stable/" \t "_blank)**“

Did you know that we currently generate 2.5 quintillion bytes of data every day? That’s a lot of generated data, and it needs to be stored, processed, and analyzed before anyone can derive meaningful information from it. Fortunately, we have Hadoop to deal with the issue of big data management.

## ****What is Hadoop?****



**Hadoop** is an open-source software framework used for storing and processing **Big Data** in a distributed manner on large clusters of commodity hardware. Hadoop is licensed under the Apache v2 license.

Hadoop is a framework that manages big data storage by means of parallel and distributed processing. Hadoop is comprised of various tools and frameworks that are dedicated to different sections of data management, like storing, processing, and analyzing. The Hadoop ecosystem covers [Hadoop](https://www.simplilearn.com/tutorials/hadoop-tutorial/what-is-hadoop" \o "Hadoop" \t "_blank) itself and various other related [big data tools](https://www.simplilearn.com/big-data-analytics-tools-article).

## ****Hadoop Ecosystem****

So far you would have figured out that Hadoop is neither a programming language nor a service, it is a platform or framework which solves Big Data problems. You can consider it as a suite which encompasses a number of services for ingesting, storing and analyzing huge data sets along with tools for configuration management.

