

## The Birth of Hadoop

Back in 2006, most traditional enterprises were still blissfully unaware of big data challenges and could not yet appreciate its opportunities. Big data was the domain of Internet giants, and one such company came to a point where it had to achieve a lot of the business outcomes we reviewed earlier. Given its scale and the size of the data sets it started to accumulate, it also had to solve this challenge in a cost-effective manner. After all, since the value of big data is typically proportional to the volume of data available for analysis, it made no sense to pay for traditional databases and/or run them on custom hardware appliances. Whatever the solution, it had to:

- Be free from draconian licensing costs
- Run on commodity hardware without requirements for custom servers and/or networking
- Scale linearly with the growth of data volume
- Afford efficient data processing and analytics that would scale well with the size of the data.

The name of the company was Yahoo! Inc. and a guy by the name of Doug Cutting, who had just joined it, had a solution in mind. A few years before, he and a friend of his, Mike Cafarella, started hacking on a project named after Doug's son's elephant toy: Hadoop.

---



## **Hadoop the Elephant**

(from the personal archive of Konstantin "Cos" Boudnik - used with permission)

The project was inspired by how another Internet giant (Google) solved similar problems back in 2000s, and aimed at clean room implementation of two seminal papers that Google published in 2003 and 2004:

- *The Google File System*
- *MapReduce: Simplified Data Processing on Large Clusters.*

It would be fair to say that, while the ideas behind Hadoop were clearly conceived at Google, Yahoo! gave us the Hadoop we know and love today.

---

