

0.1 Introduction

Before the explosion of the internet there was little need for distributed computing, most heavy processing was done by mainframe solutions. In the late 90's there came a huge surge in internet user activity, more and more people were purchasing personal computers and getting connected.

In the early 2000's there came a need to process more and more data. This data was so massive and so unstructured that nothing at the time could process it. This kind of data became known as Big Data. In 2005 Hadoop was created by Doug Cutting and Mike Cafarella to support distribution for the nutch search engine at Yahoo.

Fast forward to 2018 and we can see that Hadoop is used by many organizations to process massive amounts of data. This in part is due to the rise of social networking. Companies including Facebook, Google and LinkedIn owe much of their success to Big Data analysis and Hadoop.

In this paper I would like to explore what Hadoop is and how it is different from traditional architectures.

0.2 Hadoop Overview

Bla