

Data in big data facilities	
Period	Amount
1990–2000	Terabytes
2000–2010	Petabytes
2010–2020	Exabytes

World wide production of data	
Period	Amount
1990–2000	Petabytes
2000–2010	Exabytes
2010–2020	Zettabytes

Table 5.1: The amount of data managed by big data facilities and the approximate amount of online data in the world.

data facilities will manage exabytes of data and the world wide production of data will be measured in zetabytes.

In practice, engineers managing big data today no longer speak in terms of terabytes and petabytes. Instead they speak in terms of megawatts (MW). For example, an engineer might say that she works at a 15 MW data center that manages hundreds of petabytes of data.

5.3 The Commoditization of Data

The amount of new data produced each year is rapidly growing, but estimating exactly how much new data is produced each year is not straightforward. A good analysis was done by a research group led by Peter Lyman and Hal Varian at the School of Information Management and Systems at the University of California at Berkeley in 2000 [81]. Although outdated now, it is still very informative. The study clearly showed that by 2000 data was becoming commoditized.

Interestingly, the study was supported by EMC, a ven-