Καρακασίδης Α. Κολωνιάρη Γ. Γούναρης Α. Παπαδόπουλος Α. Ν.

Τεχνολογίες Επεξεργασίας και Ανάλυσης Μεγάλων Δεδομένων

METADATA

● ● ●

Abstract

This book comes to fill a large gap that exists in the Greek literature regarding the processing and analysis of big data. More specifically, it aims to present and use the most commonly used systems and techniques for managing and extracting useful knowledge from data characterized by large volume, potentially fast renewal rates and diversity in terms of their structure. In more detail, in this book we study specific application development methodologies as well as specific systems. For example, we study the MapReduce model and how it is implemented in Hadoop, Spark, and other systems. In addition, the Hadoop ecosystem is studied in detail, as well as the Spark application development methodology, while examples are given in Scala and Python. Also, NOSQL systems are presented, emphasizing the four different categories of systems, depending on their functionality. In the book we also study techniques for data stream management and how it is carried out by the systems, while in a separate chapter we study issues of graph data processing and analysis. Then, the basics of managing resources in a distributed system using different resource management systems and how the applications are run are reviewed. Also, in a separate chapter we study ways of processing data in GPU systems that are also widely used for data analysis in a parallel way. Installation details are also given in the book. The book is addressed to postgraduate students and mainly to courses directly related to topics related to the processing and analysis of big data using modern distributed and parallel systems. However, part of the book could work as a supplement in undergraduate courses with a similar subject.