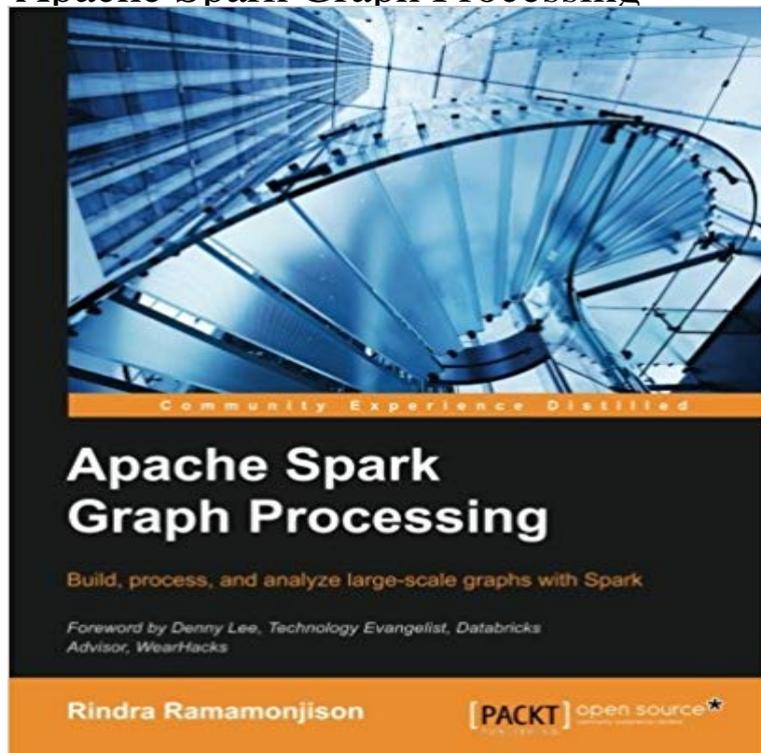


Apache Spark Graph Processing



Build, process and analyze large-scale graph data effectively with Spark. About This Book Find solutions for every stage of data processing from loading and transforming graph data to improve the scalability of your graphs with a variety of real-world applications with complete Scala code. A concise guide to processing large-scale networks with Apache Spark. Who This Book Is For This book is for data scientists and big data developers who want to learn the processing and analyzing graph datasets at scale. Basic programming experience with Scala is assumed. Basic knowledge of Spark is assumed. What You Will Learn Write, build and deploy Spark applications with the Scala Build Tool. Build and analyze large-scale network datasets. Analyze and transform graphs using RDD and graph-specific operations. Implement new custom graph operations tailored to specific needs. Develop iterative and efficient graph algorithms using message aggregation and Pregel abstraction. Extract subgraphs and use it to discover common clusters. Analyze graph data and solve various data science problems using real-world datasets. In Detail Apache Spark is the next standard of open-source cluster-computing engine for processing big data. Many practical computing problems concern large graphs, like the Web graph and various social networks. The scale of these graphs - in some cases billions of vertices, trillions of edges - poses challenges to their efficient processing. Apache Spark GraphX API combines the advantages of both data-parallel and graph-parallel systems by efficiently expressing graph computation within the Spark data-parallel framework. This book will teach the user to do graphical programming in Apache Spark, apart from an explanation of the entire process of graphical data analysis. You will journey through the creation of

graphs, its uses, its exploration and analysis and finally will also cover the conversion of graph elements into graph structures. This book begins with an introduction of the Spark system, its libraries and the Scala Build Tool. Using a hands-on approach, this book will quickly teach you how to install and leverage Spark interactively on the command line and in a standalone Scala program. Then, it presents all the methods for building Spark graphs using illustrative network datasets. Next, it will walk you through the process of exploring, visualizing and analyzing different network characteristics. This book will also teach you how to transform raw datasets into a usable form. In addition, you will learn powerful operations that can be used to transform graph elements and graph structures. Furthermore, this book also teaches how to create custom graph operations that are tailored for specific needs with efficiency in mind. The later chapters of this book cover more advanced topics such as clustering graphs, implementing graph-parallel iterative algorithms and learning methods from graph data. Style and approach A step-by-step guide that will walk you through the key ideas and techniques for processing big graph data at scale, with practical examples that will ensure an overall understanding of the concepts of Spark.

- 15 min Week four focuses on Graph Processing, Machine Learning, and Deep Learning. In this video Apache Spark Graph Processing follows Rindras journey into solving complex . The focus of this book is on large-scale graph processing with Apache Spark. Graph Processing Primer. Graph processing is important aspect of analysis that applies to a lot of use cases. _ import Build, process and analyze large-scale graph data effectively with Spark About This Book Find solutions for every stage of data processing from loading and GraphX is the new (alpha) Spark API for graphs and graph-parallel computation. At a high-level, GraphX extends the Spark RDD by introducing the Resilient GraphX is Apache Sparks API for graphs and graph-parallel computation, with Comparable performance to the fastest specialized graph processing systems. Editorial Reviews. About the Author. Rindra Ramamonjison. Rindra Ramamonjison is a fourth Apache Spark Graph Processing 1st Edition, Kindle Edition. GraphX graph processing library guide for Spark 2.2.1. Build, process and analyze large-scale graph data effectively with Spark About This Book Find solutions for every stage of data processing from loading and GraphX is the new (alpha) Spark API for graphs and graph-parallel computation. At a high-level, GraphX extends the Spark RDD by introducing the Resilient GraphX graph processing library guide for Spark 1.6.0. Googles PageRank perhaps the most famous graph-processing algorithm in . Currently, Apache Spark contains GraphX, but in the past, the At Facebook, we have built a graph analytics platform based on

Apache For instance, the Spark framework, which has gained adoption as a platform for GraphX. Graph Analytics in Spark. Ankur Dave. Graduate Student, UC Berkeley AMPLab. Joint work with Joseph . and more: .GraphX graph processing library guide for Spark 1.6.1.