PRENTICE HALL

UNDERSTANDING BIG DATA SCALABILITY

BIG DATA SCALABILITY SERIES, **PART I**

CORY ISAACSON

PRAISE FOR *UNDERSTANDING BIG DATA SCALABILITY*

"This book is useful to anyone who works with data and wants to learn more about scaling. Cory helps you understand what causes databases to slow down as data volumes grow over time. He then reviews a number of strategies that you have at your disposal to manage the growth, including software and database tuning, hardware upgrades, read-replication, and ultimately horizontal partitioning of data."

—Dan Lynn, cofounder of FullContact

"*Understanding Big Data Scalability* presents the fundamentals of scaling databases from a single node to large clusters. It provides a practical explanation of what Big Data systems are, and the fundamental issues to consider when optimizing for performance and scalability. Cory draws on his many years of database experience to explain the issues involved in working with data sets that can no longer be handled with single monolithic relational databases.

"When transitioning from a traditional relational database deployment, it is tempting to ignore traditional database discipline regarding data modeling and data integrity. Much of this has been motivated by the proliferation of schema-less NoSQL databases. In spite of this trend, Cory shows why it is still important to carefully structure your data to maintain data integrity and allow sharding in such a way as to avoid costly distributed scan/shuffle operations. He discusses a practical approach to this called *relational sharding*. This is a commonsense method that avoids the pitfalls of black-box sharding. Cory's approach is particularly relevant now that relational data models are making a comeback via SQL interfaces to popular NoSQL databases and Hadoop distributions.

"*Understanding Big Data Scalability* addresses practical problems in Big Data processing systems using real-life examples. This book should be especially useful to database practitioners new to the process of scaling a database beyond a traditional single-node deployment."

—Brian O'Krafka, software architect

Understanding Big Data Scalability: Big Data Scalability Series, Part I

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents