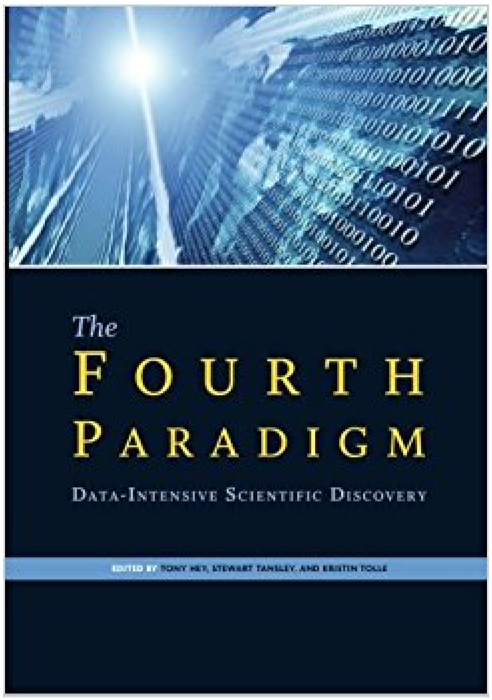
The Fourth Paradigm: Data-Intensive Scientific Discovery



Synopsis

This book presents the first broad look at the rapidly emerging field of data-intensive science, with the goal of influencing the worldwide scientific and computing research communities and inspiring the next generation of scientists. Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets. The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud-computing technologies. This collection of essays expands on the vision of pioneering computer scientist Jim Gray for a new, fourth paradigm of discovery based on data-intensive science and offers insights into how it can be fully realized.

Book Information

Paperback: 284 pages

Publisher: Microsoft Research; 1 edition (October 16, 2009)

Language: English

ISBN-10: 0982544200

ISBN-13: 978-0982544204

Product Dimensions: 7 x 0.7 x 10 inches

Shipping Weight: 1.4 pounds (View shipping rates and policies)

Average Customer Review: 4.0 out of 5 stars 13 customer reviews

Best Sellers Rank: #574,337 in Books (See Top 100 in Books) #98 inÂ Books > Medical Books > Medical Informatics #214 inÂ Books > Computers & Technology > Networking & Cloud Computing > Network Administration > Storage & Retrieval #542 inÂ Books > Science & Math > Experiments, Instruments & Measurement > Methodology & Statistics

Customer Reviews

Tony Hey is Vice President of Microsoft Research and has published over 150 research papers on particle physics, computer science and multi-disciplinary eScience. He is the co-author, with Ian Aitchison, of the graduate textbook Gauge Theories in Particle Physics (4th edition) and co-editor of Richard Feynman’s Lectures on Computation. He is also co-author of popular science books on quantum mechanics and relativity, The New Quantum Universe and Einstein’s Mirror. Most recently, he co-edited The Fourth Paradigm, a book about eScience and the new era of data-intensive science.

"We have to do better producing tools to support the whole research cycle - from data capture and data curation to data analysis and data visualization." - Jim GrayThe Fourth Paradigm is a collection of papers talks on research areas that aim to improve the research cycle. The talks are a memorial to Microsoft Tech Fellow Jim Gray. Gray had the insight that science has gone through four paradigms so far. The first paradigm, which has lasted over the last few thousand years, was empirical science which describes natural phenomena. Over the last few hindered years, the second paradigm of theoretical science using models and generalizations has occurred. Within the last 50 to 70 years, the third paradigm of computational science has developed to simulate complex phenomena. Finally, the fourth paradigm (also known as eScience) has developed to unify theory, experiment, and simulation. Jim Gray says:"... it is worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration."The book itself is divided into four major sections: Earth and Environment, Health and Wellbeing, Scientific Infrastructure, and Scholarly Communications with 6 to 8 papers per section. The emphasis here is on science; however, I'd assert that all these areas directly impact engineering as well. For example, the flight test of a new product involves an enormous amount of data, which produces much analysis, knowledge, and understanding. The principle idea of eScience (and eEngineering) is that the data and analysis interoperate with each other, such that information is at your fingertips for everyone, everywhere. The payoff is a large increase in information velocity and productivity. In the end, an analysis or report will be an overlay on the data. I have seen this start to happen, and agree with Jim Gray that our current tools are very primitive - a lot of new tools are going to be required.A paper that I found particularly interesting was "Discovering the Wiring of the Brain" Their summary is: "Decoding the complete connectome of the human brain is one of the great challenges of the 21st century." I agree - and discovering the scientific and engineering applications that will emerge will be even more of a challenge. This is an area that requires an entire new way to handle all the data - consider that a snapshot 1 cubic mm of image data from a human brain contains a petabyte of data, and that a human brain contains about one million cubic mm.This fascinating book is availed for free download at the Microsoft Research website at [...]/I think any working scientist or engineer will find much to learn and think about in this collection of papers on the emerging Fourth Paradigm and the world of eScience (and eEnginnering)

I am a fanatic of data, technology and enterprise. This cover the main tendencies and uses of scientific computing, big data, etc. and their challenges, touching interesting points to be considered and some drivers of the change. There are much more opportunities to come.

This was a much referenced book. It proved to be insightful in many areas but it also tended to be very relational database focused. It missed a lot of what is going on in object databases, semantics, very very large data structures. Scale is one of the issues and this broached the subject but only discussed it in terms of relational models. One might argue that relational models are good for reductionist efforts but synthesis efforts fail because they become compute bound.

I really enjoy the subject matter: science and computers.However, all of the papers were top-down overviews. I wanted to dig into some case studies. For example, Microsoft has a working project: World Wide Telescope. How many data sources do they use? How do they blend data from conflicting sources? How do they curate the data? How much telescope gear (and how much computer hardware and software) would I need to contribute? None of the essays went into details on these projects.Several papers did make some useful, interesting points.Much of scientific research today is cottage industry: one group puts together some instruments, gathers data, analyzes it, publishes paper. A revolution akin to the industrial revolution will happen: specialized groups will operate instruments and publish data; other groups will analyze the data.The data repositories of the future must accommodate large numbers of disparate groups gathering data -- and the scientific community must reward them. Data organization, provision of metadata, provenance are all big unsolved questions. (I’d have liked more detailed information here, too).Some scientific instruments collect data so fast that the bottleneck is no longer data acquisition but data interpretation. Similarly, data repositories are so large that making copies of the dataset is expensive -- it will actually be cheaper for data repositories to offer services where researchers run custom programs against the data.This high-level overview is grand, but it’s hard to test. Surely these pronouncements are based on experience in actual scientific projects. I wanted to read more at this lower level.

Very compelling read.

Seems more like business side book !!!

Very timely topic that is explained lucidly for the layman and expert alike.The book presents the issues in a comprehensive and interesting fashion.

good quality!

The Fourth Paradigm: Data-Intensive Scientific Discovery Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data Book 1) Data Analytics: Applicable Data Analysis to Advance Any Business Using the Power of Data Driven Analytics (Big Data Analytics, Data Science, Business Intelligence Book 6) Big Data For Business: Your Comprehensive Guide to Understand Data Science, Data Analytics and Data Mining to Boost More Growth and Improve Business - Data Analytics Book, Series 2 Discovery Map 85: Cork Kerry (Discovery Maps): Cork Kerry (Discovery Maps) (Irish Discovery Series) Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) Routledge Intensive Russian Course (Routledge Intensive Language Courses) Core Curriculum for Neonatal Intensive Care Nursing, 5e (Core Curriculum for Neonatal Intensive Care Nursing (AWHONN)) Certification and Core Review for Neonatal Intensive Care Nursing, 4e (Watson, Certification and Core Review for Neonatal Intensive Care Nursing) Core Curriculum for Neonatal Intensive Care Nursing, 4e (Core Curriculum for Neonatal Intensive Care Nursing (AWHONN)) Data Classification: Algorithms and Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series) Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data Analytics: Business Intelligence, Algorithms and Statistical Analysis (Predictive Analytics, Data Visualization, Data Analytics, Business Analytics, Decision Analysis, Big Data, Statistical Analysis) Data Analytics and Python Programming: 2 Bundle Manuscript: Beginners Guide to Learn Data Analytics, Predictive Analytics and Data Science with Python Programming Data Analytics For Beginners: Your Ultimate Guide To Learn and Master Data Analysis. Get Your Business Intelligence Right â “ Accelerate Growth and Close More Sales (Data Analytics Book Series) Discovering Knowledge in Data: An Introduction to Data Mining (Wiley Series on Methods and Applications in Data Mining) Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems) Transforming Mission: Paradigm Shifts in Theology of Mission (American Society of Missiology) The Cannabis Manifesto: A New Paradigm for Wellness