

Promoting Skills for the Data Driven Economy

Devdatt Dubhashi

Dept. of Computer Science and Engg.

Chalmers and Gothenburg University

Sweden

Data Intensive Science

ABOUT THE FOURTH PARADIGM

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.

"The impact of Jim Gray's thinking is continuing to get people to think in a new way about how data and software are redefining what it means to do science."

—BILL GATES

"I often tell people working in eScience that they aren't in this field because they are visionaries or super-intelligent—it's because they care about science and they are alive now. It is about technology changing the world, and science taking advantage of it, to do more and do better."

—RHYS FRANCIS, AUSTRALIAN RESEARCH INFRASTRUCTURE COUNCIL

"One of the greatest challenges for 21st-century science is how we respond to this new era of data-intensive science. This is recognized as a new paradigm beyond experimental and theoretical research and computer simulations of natural phenomena—one that requires new tools, techniques, and ways of working."

—DOUGLAS KELL, UNIVERSITY OF MANCHESTER

"The contributing authors in this volume have done an extraordinary job of helping to refine an understanding of this new paradigm from a variety of disciplinary perspectives."

—GORDON BILL, MICROSOFT RESEARCH

THE FOURTH PARADIGM



The FOURTH PARADIGM

DATA-INTENSIVE SCIENTIFIC DISCOVERY

HEY
TANSLEY
TOLLE

EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

01010101



Microsoft
Research

The Second Machine Age

- “**Bounty**” of ICT and AI (Auto cars, natural language technology ...)
- “**Spread**” : uneven distribution of gains
- ICT as a **general purpose tech.** and other technologies

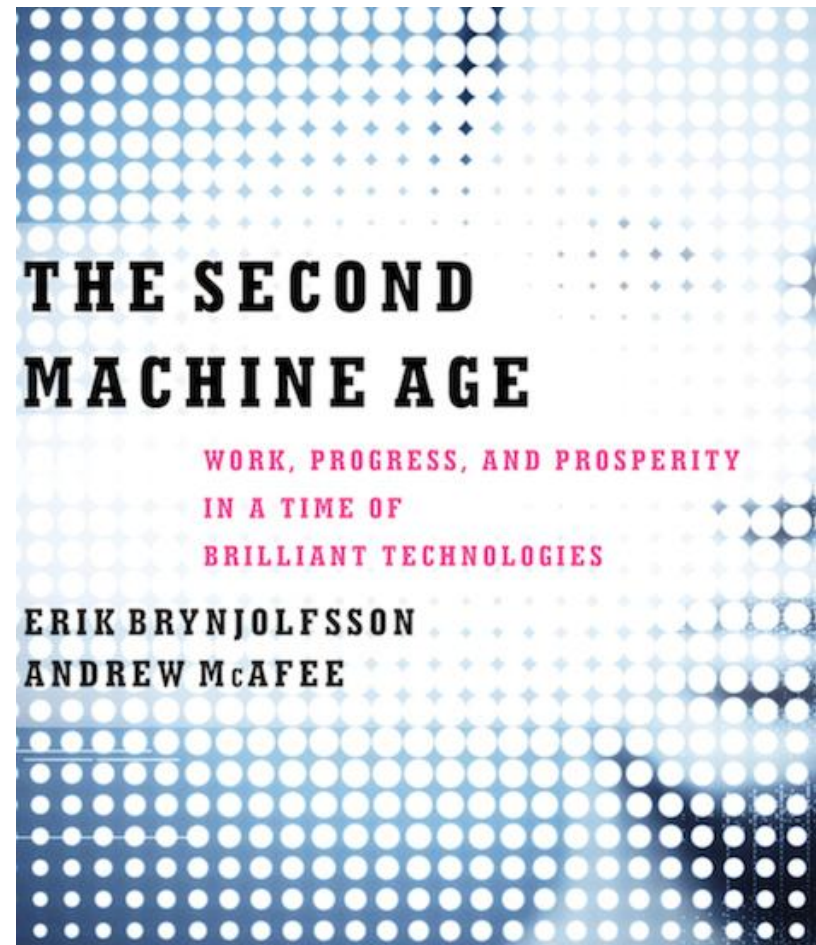


Image and Intervention Center Sahlgrenska

- Vastra Gotalands Region
- Sahlgrenska Hospital
- Sahlgrenska Academy
- Chalmers
- Gothenburg University



Business Intelligence: the Internet as Crystal Ball



Recorded Future

CREATING AN INSIGHTFUL WORLD

FINDWISE

SEARCH DRIVEN SOLUTIONS

Data Science

- Foundational Research
- Technologies
- Application Areas

Data Science: Foundations

- **Probability and Statistics**: Graphical models, inference, topic models, non-parametric statistics ...
- **Optimization**: convex optimization
- **Parallel and distributed algorithms**: ADMM, GraphLab

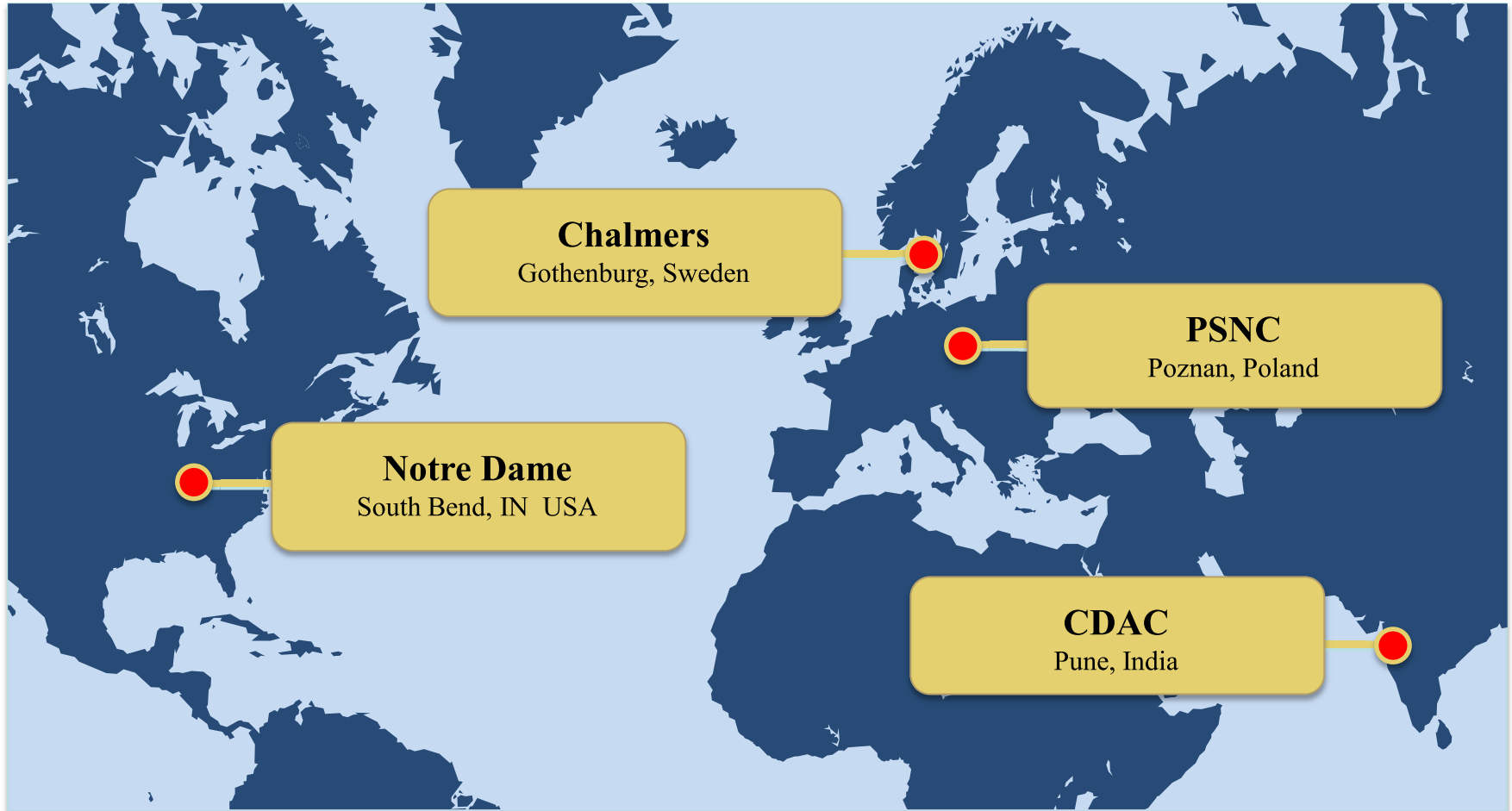
Data Science: Technologies

- **Non-structured databases:** MongoDB, Hbase, Cassandra
- **Hadoop** technologies
- **Cloud** technologies
- **High Performance Computing**

Data Science: Applications

- Life Sciences
- Business Intelligence
- Natural Language technologies
- Highenergy physics
- Transport; automated vehicles
- Energy: smart grids
- Urban Futures; smart cities
- ...

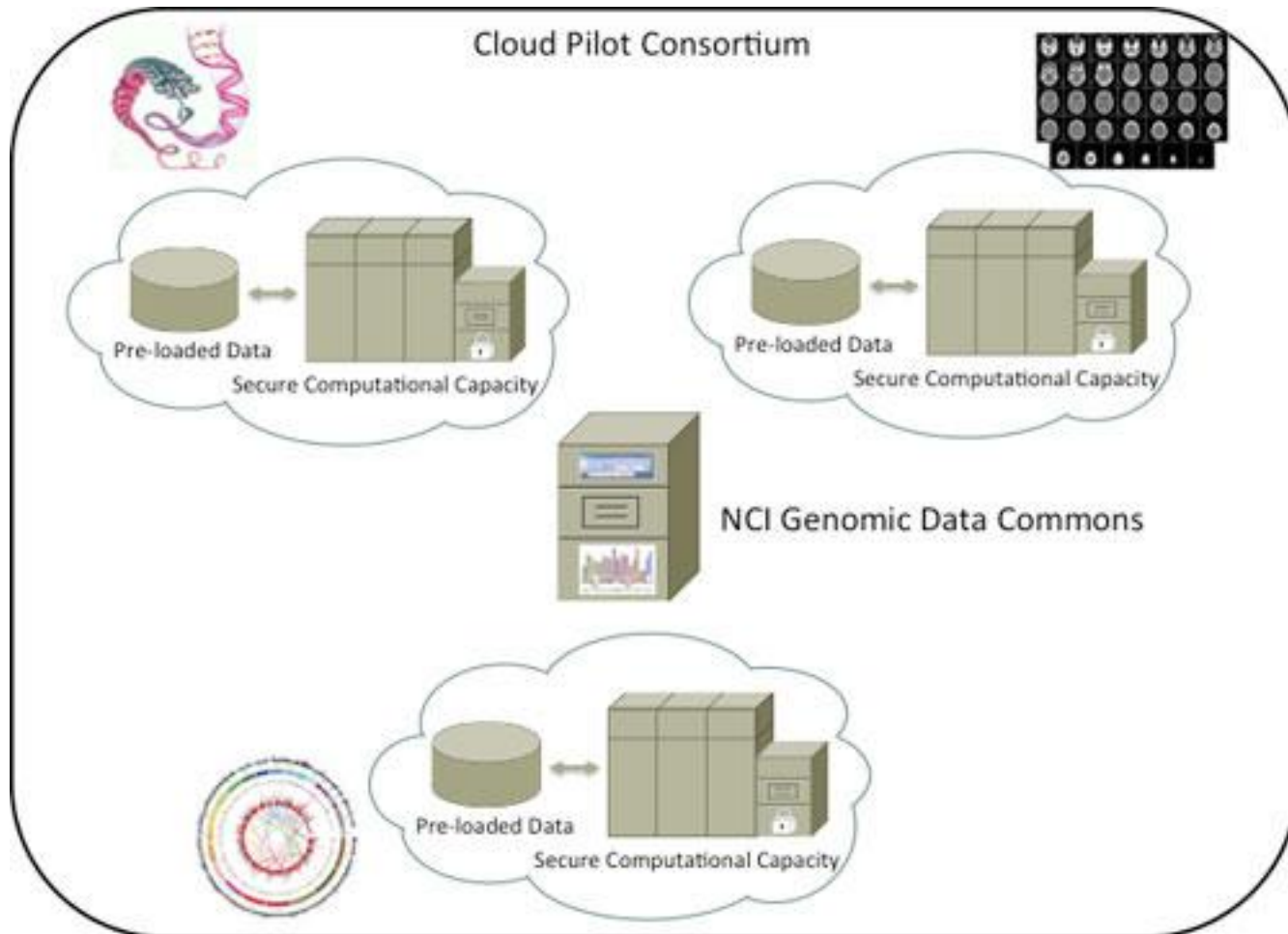
ICTBioMed – a Consortium of HPC Centers



Public-Private Consortia

- **Academia** : foundational research
- **Industry**: engineering and commercialization.

ICTBioMed Cloud



The Second Machine Age