

# Merge or Split: Mutual Influence between Big Data and HPC Techniques

---

HPBDC 2016 panel

**Panel moderator: Dr. Jianfeng Zhan**

*Professor, ICT, Chinese Academy of Sciences  
and University of Chinese Academy of Sciences*

May 27, 2016  
Chicago, USA



中国科学院  
INSTITUTE OF COMPUTING TECHNOLOGY

# Panelists

- [Chaitanya Baru](#), San Diego Supercomputer Center
- [Pete Beckman](#), Argonne National Laboratory, The University of Chicago
- [Andrew A. Chien](#), The University of Chicago, Argonne National Laboratory
- [Geoffrey C. Fox](#), Indiana University Bloomington
- [D. K. Panda](#), The Ohio State University

# Big Data vs. HPC

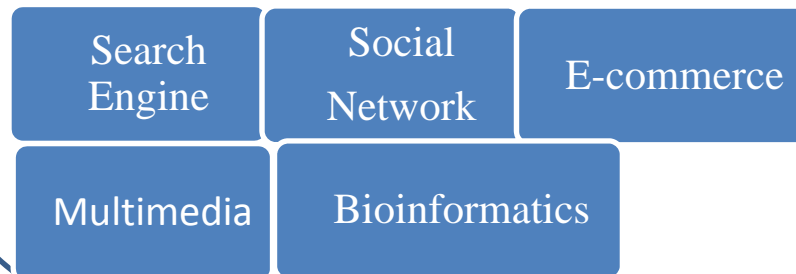
- Not well defined
  - Broadly include diversity of applications.
  - HPC
    - SC: GPGPU, IB
    - Internet services: distributed computing.
- Different perspectives
  - Computing-centric: HPC
  - Data-centric: Big data
- Benchmarking defines playground.

# BigDataBench 3.2 summary

## BDGS(Big Data Generator Suite) for scalable data

|                         |                            |                              |
|-------------------------|----------------------------|------------------------------|
| Wikipedia Entries       | Amazon Movie Reviews       | Google Web Graph             |
| Facebook Social Network | E-commerce Transaction     | ProfSearch Resumes           |
| ImageNet                | English broadcasting audio | DVD Input Streams            |
| Image scene             | Genome sequence data       | Assembly of the human genome |
| SoGou Data              | MNIST                      | MovieLens Dataset            |

## 15 Real-world Data Sets



## 33 Workloads



# Big data dwarfs

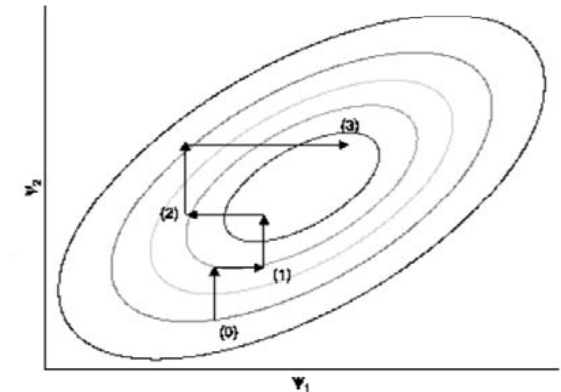
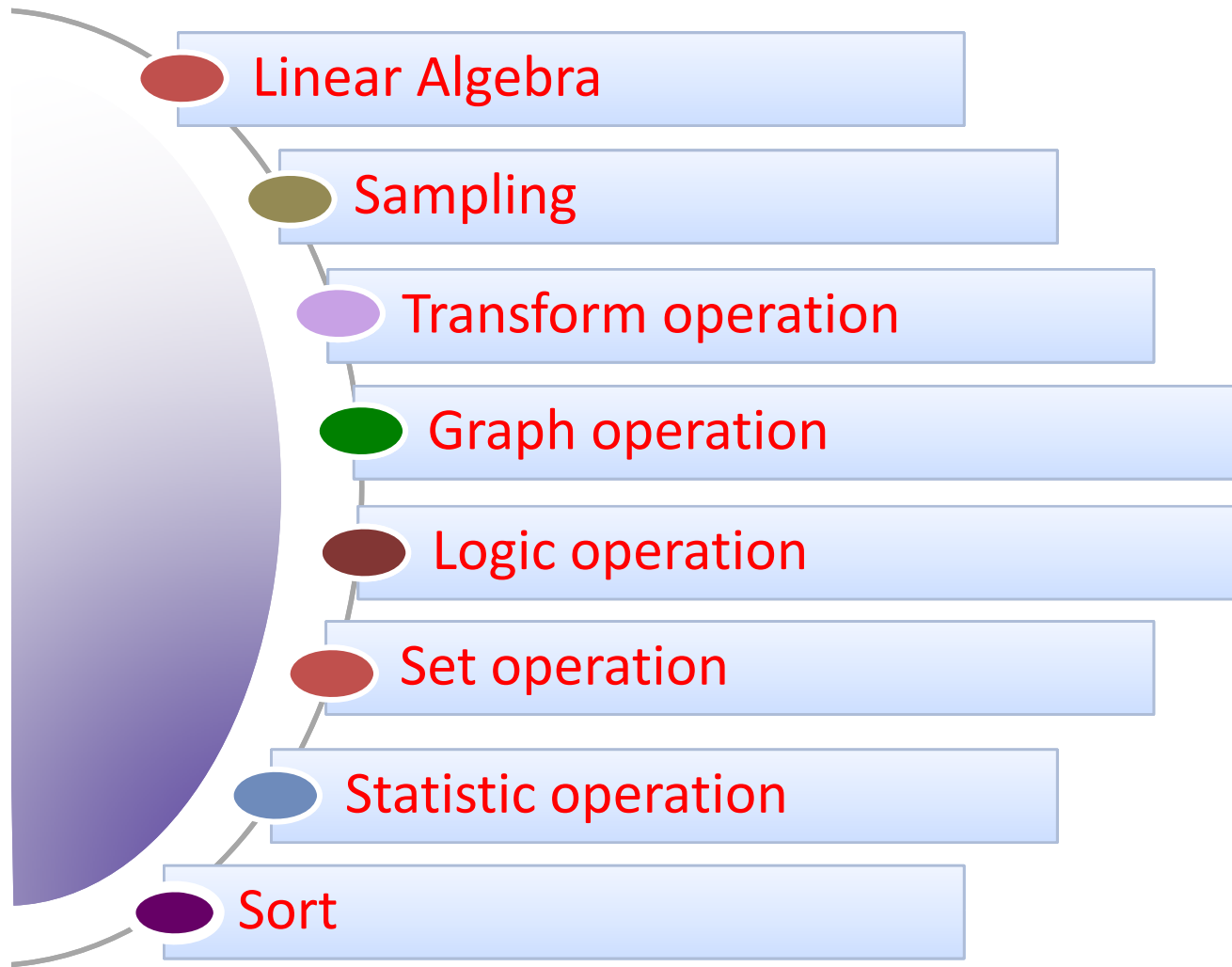
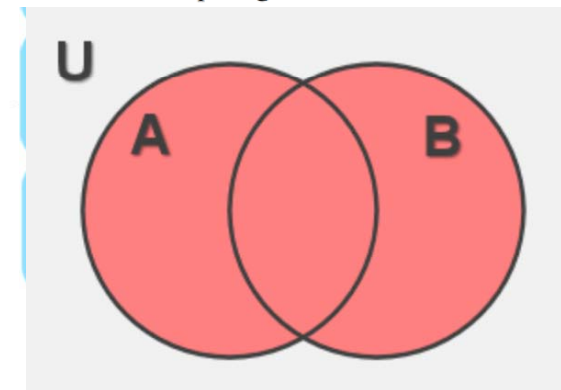


Figure 3.4: Gibbs sampling algorithm in two dimensions starting from an initial point and then completing three iterations



# Panel topics: Merge or Split?

## ■ Mutual Influence between Big Data and HPC Techniques

- What is the impact of Big Data techniques on HPC?
- What is the impact of HPC techniques on Big Data?
- Future mutual influence between HPC and Big Data techniques?

# Merge examples

- Deep learning
  - GPGPU, IB
- Many HPC applications produce big data
  - high-energy physics, astronomy, bioinformatics
- HiBD project
  - <http://hibd.cse.ohio-state.edu/>

# BigDataBench updates & trends

- Will release deep learning workloads soon
- Will include scientific computing big data workloads
  - High-energy physics
  - Astronomy
- BigDataBench project homepage
  - <http://prof.ict.ac.cn/BigDataBench>
- Gao et al. (2015). Identifying Dwarfs Workloads in Big Data Analytics. *arXiv preprint arXiv:1505.06872*.



# Why split now?

- Different concerns from scientific computing & Internet services communities
  - Revenue
    - Science needs no business revenue.
  - Cost
    - People cost
    - New technology cost.

# Schedules

- Positions from the panelists (6 minutes)
- First round of rebuttals (2 minutes)
  - Further clarify your position
  - Explicitly object to other one's position.
- Second round of rebuttals (2 minutes)
  - Argue with each other
  - Defend yourself
- Questions from the audiences



# QUESTIONS And Answers