北京航空航天大学计算机学院
Beihang University, School of Computer Science and Engineering

Welcome to Beihang U.
3rd JTC1 SGBD Meeting

Beihang University
Chunming Hu (hucm@buaa.edu.cn)

3rd JTC 1 SGBD Meeting
June 16, 2014. Beijing, China.

Welcome to Beihang University

- Beihang Univesity (1952-)
  - Research-oriented university , focusing on Aeronautics & Astronautics.
  - Information Technology is another strengthen part.

STATISTIC DATA	
Total	27811
Graduates	14428
Undergraduates	12715
Other students	668

- You may find more at :
  - http://www.buaa.edu.cn
  - http://ev.buaa.edu.cn [English Version]

The Institute of Advanced Computing Technology

SCSE@Beihang: School of Computer Science & Engineering

- History
  - 1978: Dept. of Computer Science & Eng.
  - 2002: School of Computer Science & Eng.
- Facts
  - 130+ Faculties (36 Profs. , 49 Associate Profs)
  - 9 Research labs (two are national key labs)
- **Top 4 in CS**, in Chinese Universities
  - Ranked by Ministry of Education, 2013
  - Right after Tsinghua University, PKU, NUDT

The Institute of Advanced Computing Technology

Big Data in Cyberspace

- Large Scale with Rapid Updates

1PB data in DVD:
~25km
1ZB=1PB×10^6

Social Networks

- **4 Micro-blogger Provider in China:**
- 800M Users, 200M tweets everyday, 20M+ Photos.

Internet Search

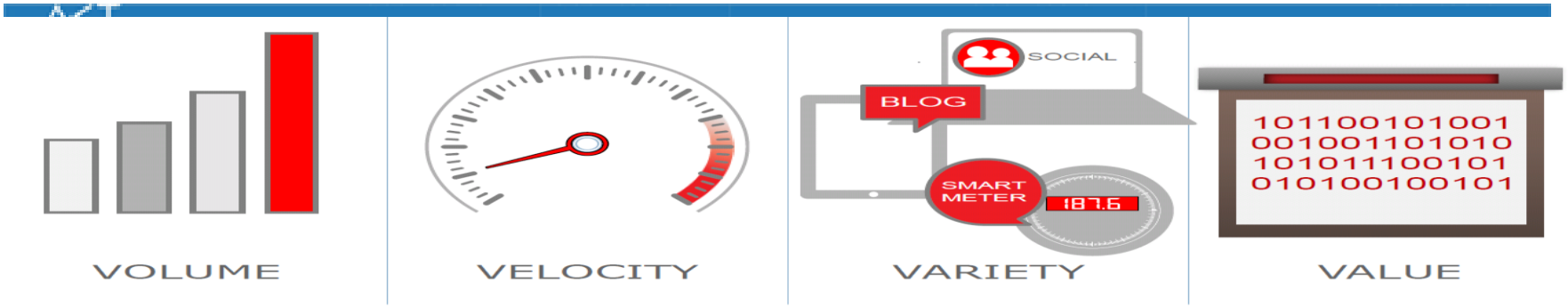
- **Baidu：** 1PB log data per Day. Handling 1000PB
- **Google：** Processing 20PB data everyday

IDC Report

- **Data doubled every 18 months**
- **Data in Cyberspace**

- **IDC Report :**
  - **2009: 0.8ZB**
  - **2012: 2.7 ZB**
  - **2020(E): 35ZB**

4V Features in Big Data

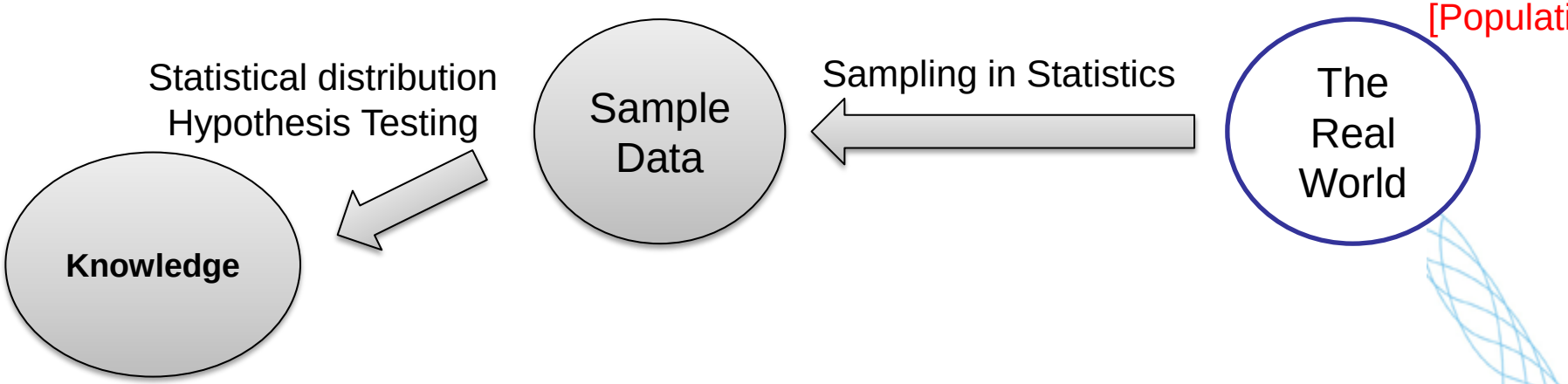


Volume	Velocity	Variety	Value
•In PB or EB •Distributed data	•Dynamic Changes •Updated constantly	•Heterogeneous •Semi-structured or unstructured	• Biz opportunity • Sensitive Data

Wikipedia

large and complex datasets, which is quite difficult to process using existing data management tools, and traditional data processing applications

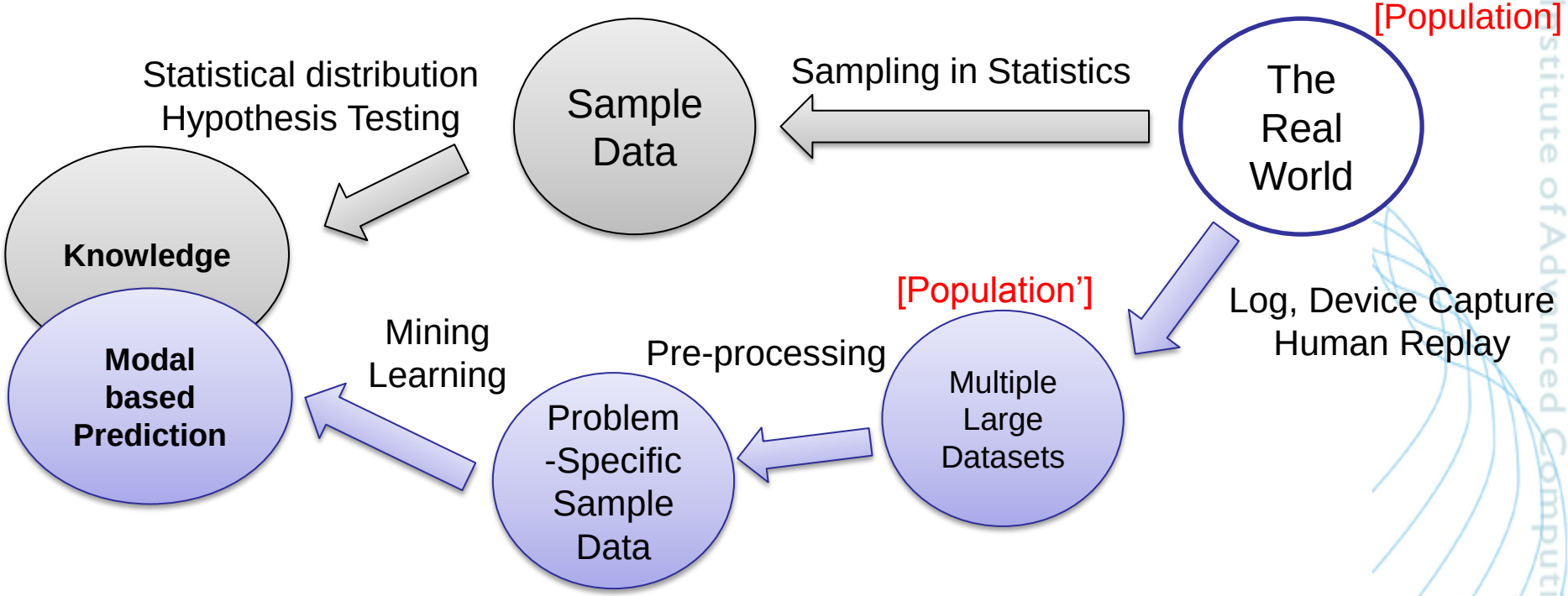
Big Data Changes



New Statistical Theory & Math Tools

The Institute of Advanced Computing Technology

Big Data Changes



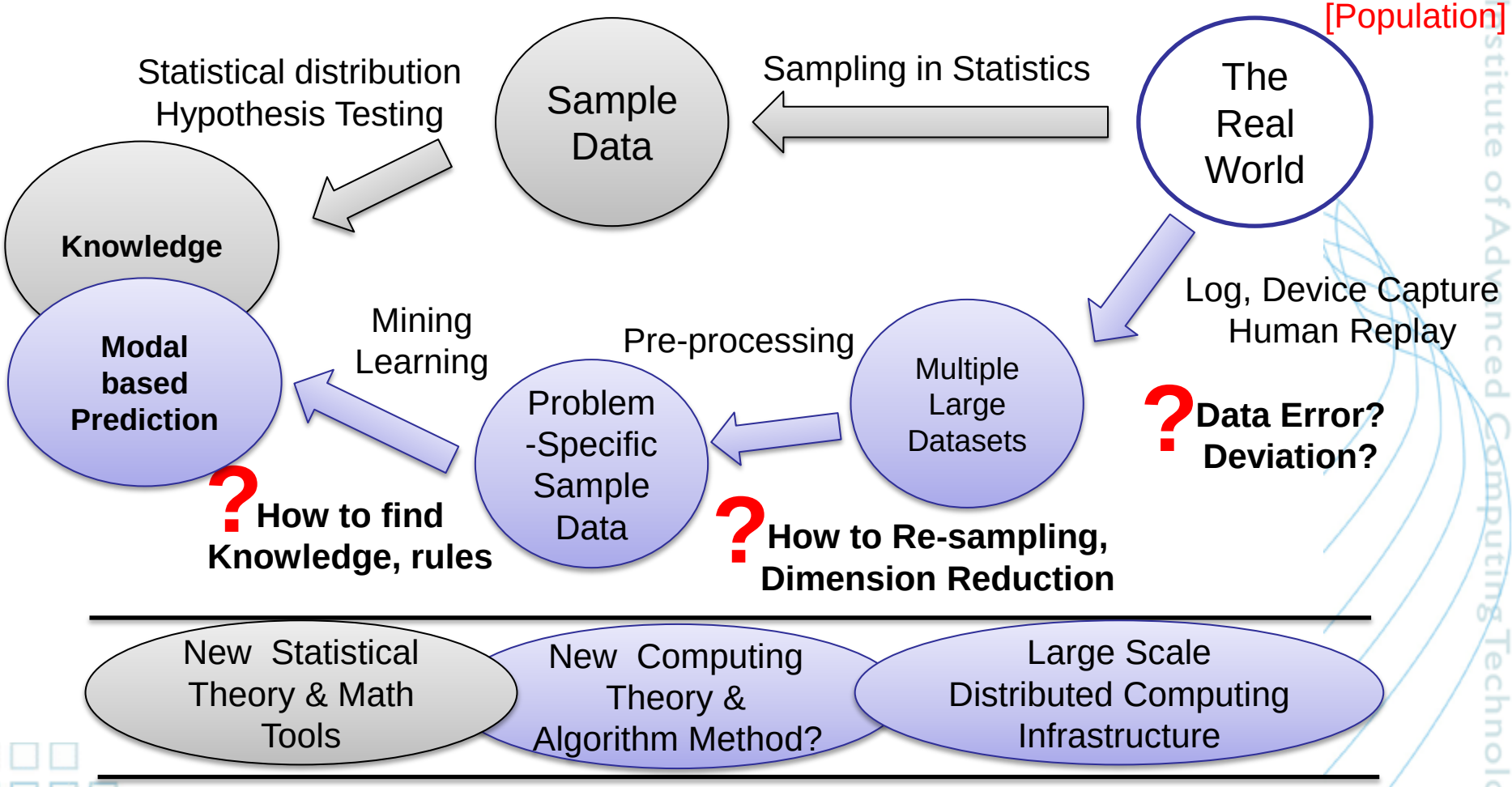
New Statistical Theory & Math Tools

New Computing Theory & Algorithm Method?

Large Scale Distributed Computing Infrastructure

The Institute of Advanced Computing Technology

Big Data Changes



The Institute of Advanced Computing Technology

Big Data Research at Beihang

软件开发环境国家重点实验室
State Key Laboratory of Software Development Environment

Mass Data Processing

1991

1993

Internet Software & Distributed Systems

2000

Computing Model and Complexity

2005

Large Scale Mobile Sensing & ITS

Digital Media Analysis

2010

Big Data Querying and Analysing

Human Centric Video Analysis

2011

Fault Tolerance Distributed System

2012

Computing Complexity in BD

Data Driven Context Generation

Theory and Algorithms | **Systems & Software** | **Domain Spec Data Processing**

2014

Big Data Computing Theory and Systems

RCBD: Research Centre on Big Data

The Institute of Advanced Computing Technology

Questions on Big Data Computing

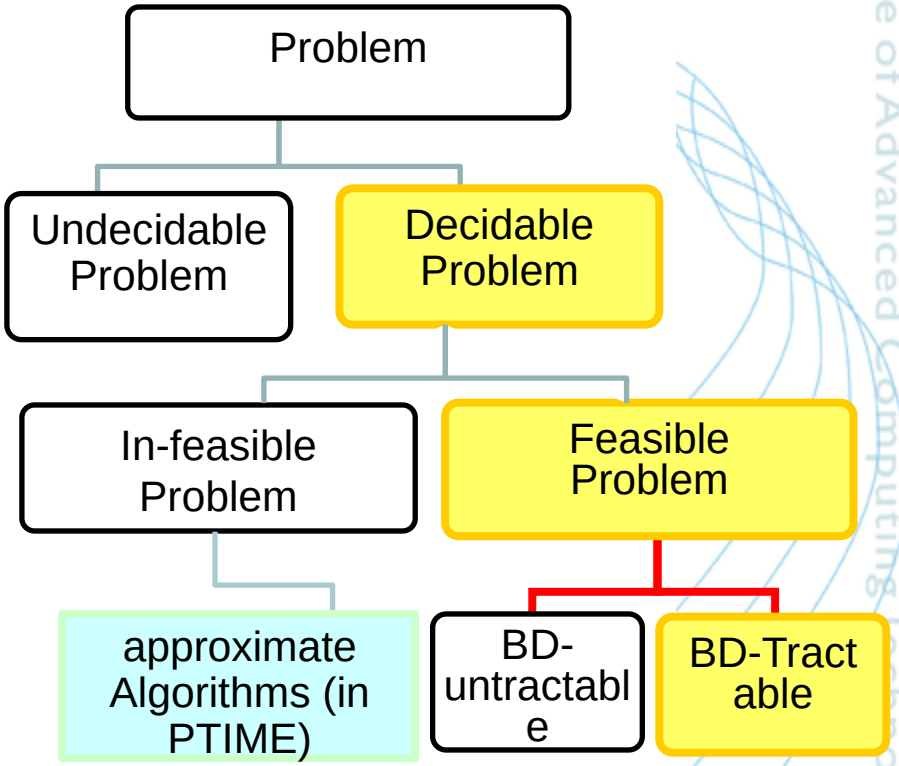
- Question 1: Is there new Theory for Big Data Computing?

$$G = F(x)$$

Algorithm

Data

  - **Good**: PTIME
  - **Bad**: NP-Hard
  - **Ugly**: PSPACE-hard, or EXPTIME-hard, undecidable



The Institute of Advanced Computing Technology

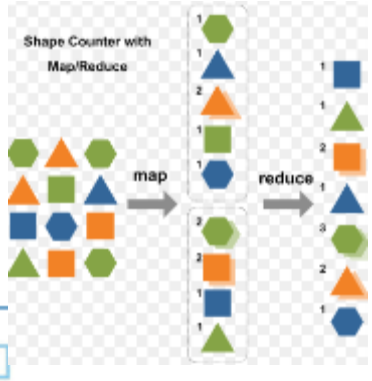
Questions on Big Data Computing

- Question 2: Is Hadoop a must to data processing?
  - Different Computing requirements, User Scenarios
  - Different Algorithm Design methods

•MapReduce (OSDI 2004)

$x \rightarrow x_1, x_2, x_3 \ldots x_n$

•Handling All Data in a Distributed way



Increm ental

3I

MR is not the only solution

•Incremental Computing: Percolator by Google (OSDI 2010)

New Algorithm Methods

• Resampling

• Query preserving data compression

• Partial evaluation and distributed processing

•Top-k query and terminating… …

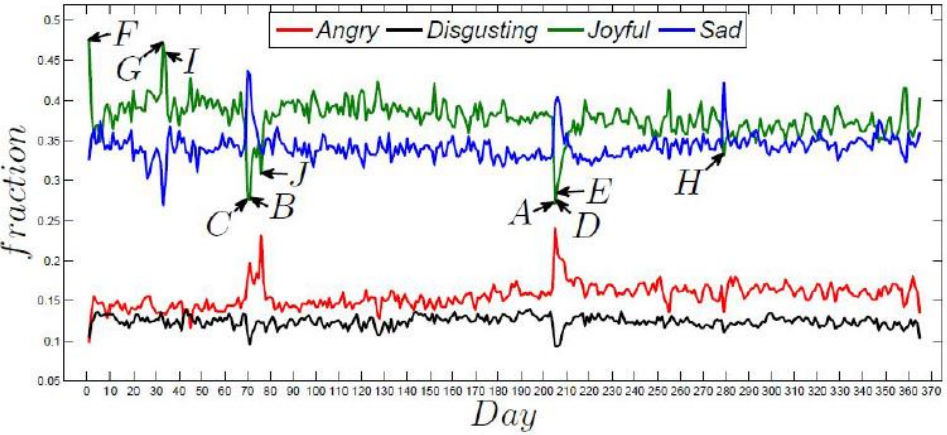
Questions on Big Data Computing

- Question 3: How to handle the data computing algorithms in an operatable manner?
  - Application/Domain specific Features Analysis
    - Data Pattern, Arriving Rate, Query, …
    - General Purpose vs. Specific Purpose
    - Domain-specific Knowledges
  - Large Scale Data Mining and ML Methods
  - Distributed system
    - Offline/Online
    - Batch/Incremental/Streaming
    - In-memory Computing

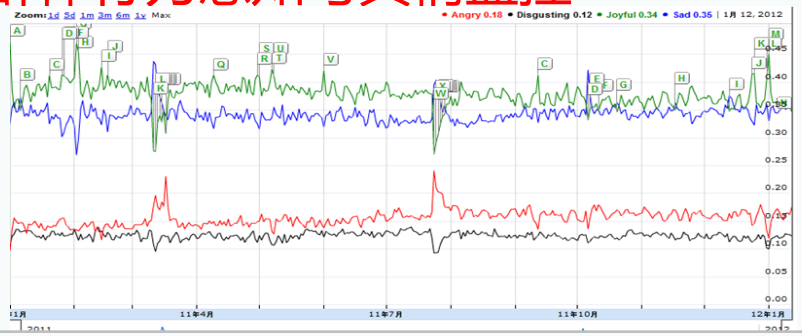
The Institute of Advanced Computing Technology

Case 1: Social Data Analysis

Sentiment	#Emoticons	Typical emoticons
Angry	*9*	
Disgusting	*14*	
Joyful	*50*	
Sad	*22*	



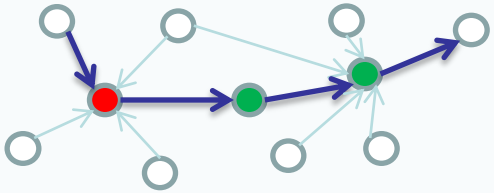
群体行为感知与舆情监控



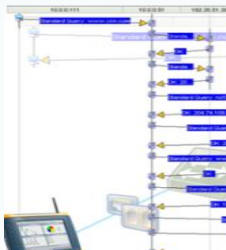
热点定位、角色分析与推荐

信息传播规律感知预测与定向干预

传播控制与疏导

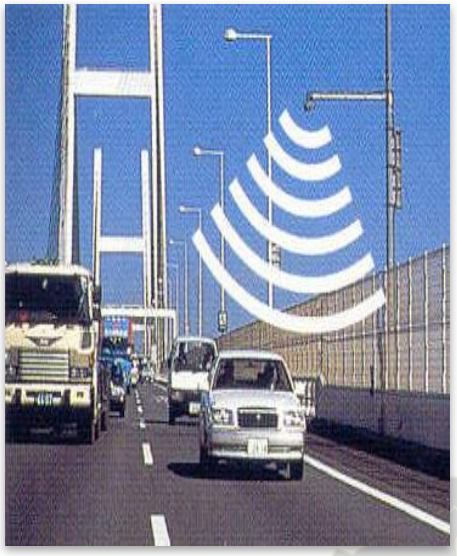


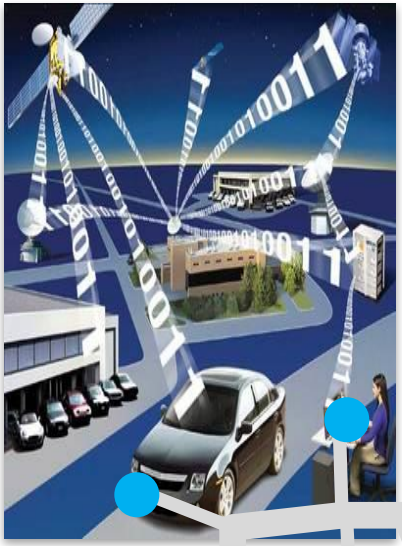
溯源与路径



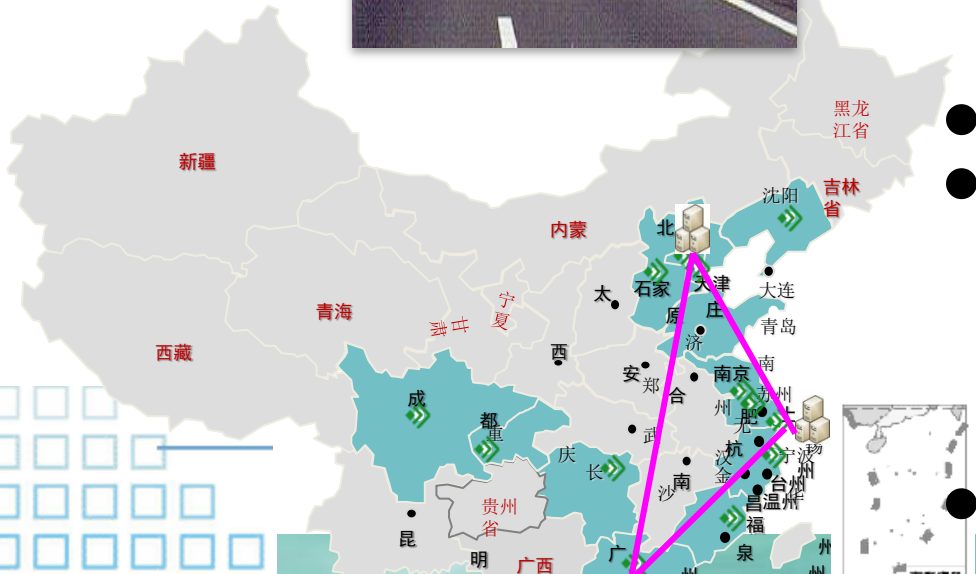
Case 2: Traffic Data Analysis

Road Facility Sensing





GPS-based Positing Sensing & Calculating



- **Euippied on 200K Vechicles**
- **Serving 40 Cities in China**
  北京、上海、广州、深圳、沈阳、宁波、成都、苏州、重庆、南京、福州、青岛、石家庄、长沙、天津……
- **Adopted by Thailand**

Case 3: Media Data Understanding

Data from CCTV

User Generated Data from Web

Geospatial Data

Big Media Data

Quality

- Media Data Enhancement

Semantics

- Media Data Semantic Mining

Tracking

- Media Object Identification & Tracking

RCBD: International Research Centre on Big Data

International Research Centre on Big Data (Founded in Sept 2012)

大数据科学与工程国际研究中心
RCBD
International Research Center on Big Data

- Beihang U.
- U. Edinburgh
- HKUST
- U.Pennsylvania
- Baidu

广州市香港科大霍英东研究院

ISCAS

HKUST Fok Ying Tung Graduate School

HUAWEI

NOAH'S ARK LAB
From Big Data to Deep Knowledge

The Institute of Advanced Computing Technology

973 Project on Big Data

- Computing Theory on Big Data (2014-2018)
  - a 973 Project Funding from Ministry of Science & Technology, with 8 institutional organization

北京航空航天大學 BEIHANG UNIVERSITY

ISCAS

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

國防科學技術大學 National University of Defense Technology

HUAWEI

- Funding: 20M RMB Yuan
- Focusing on the Computing Issue of Big Data Processing

CNCERT/CC 国家计算机网络应急技术处理协调中心

Baidu 百度

The Institute of Advanced Computing Technology

W3C China Host

- **W3C China Host (www.w3.org)**
  - Running China office since 2006, has become host in 2013
  - 36 Chinese members, bridges to Internet Industry
  - http://www.chinaw3c.org [Chinese]

The Institute of Advanced Computing Technology

Data Activity in W3C

- **W3C is aware of (Big) Data**
  - Nov. 2013: Information & Knowledge Domain (INK)
  - **Data Formats**: XML, CSV on the Web, Geospatial Data…
  - **Data Semantics**: RDFa, RDF Data Shapes
  - **Data APIs for Web Apps**: Web Storage
  - **Linked Data Platform**: LDP, e-Government
  - **Data Security & Privacy**: Privacy Protection
  - **Use Cases**: Data on the Web Best Practice
- Industry Interests: Baidu, Tencent, …

The Institute of Advanced Computing Technology

Summary & Suggestions

- Open to join/contribute the discussion of Big Data Standard research

- Use Case Study
  - Social data, Traffic data, Video data
- Big Data Architecture
  - Distributed Systems and Software
- Data on the Web
  - What may data do to build better OWP

The Institute of Advanced Computing Technology

北京航空航天大学计算机学院
Beihang University, School of Computer Science and Engineering

Thank You & Enjoy Stay in Beijing :-)

School of Computer Science, Beihang University
Chunming Hu (hucm@buaa.edu.cn)

3rd JTC 1 SGBD Meeting
June 16, 2014. Beijing, China.

Useful Links

- Beihang University, http://www.buaa.edu.cn
- A Public Video for Beihang University, http://ev.buaa.edu.cn/study_at_buaa/education/publicity_film/77351.htm
- School of Computer Science & Engineering http://scse.buaa.edu.cn
- RCBD, International Research Centre of Big Data, Beihang University, http://rcbd.buaa.edu.cn/zh/index.html
- World Wide Web Consortium (W3C), http://www.w3.org, http://www.chinaw3c.org
- W3C Information & Knowledge Domain, http://www.w3.org/INK/
- W3C Data Activity, http://www.w3.org/2013/data/Activity.html
- ISO/IEC JTC 1 Study Group on Big Data (SGBD) , http://jtc1bigdatasg.nist.gov/home.php
- 3rd JTC 1 SGBD Meeting, http://jtc1bigdatasg.nist.gov/programAsia.php

The Institute of Advanced Computing Technology