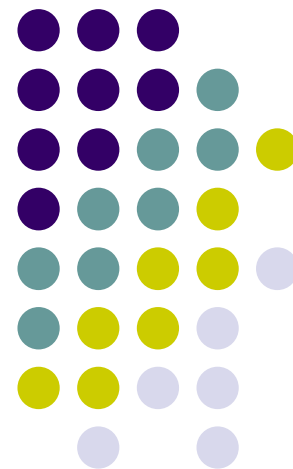


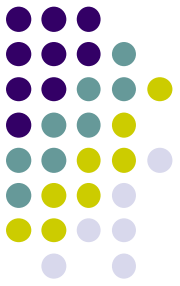
# Introduction to Computer Science

---

Jiaheng Lu  
Department of Computer Science  
Renmin University of China  
[www.jiahenglu.net](http://www.jiahenglu.net)

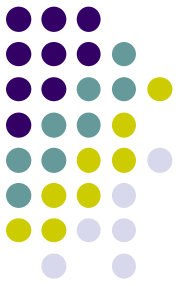


# Survey questions (19 students)



- ( ) Introduction to computer science
- ( ) Introduction to computer scientists
- ( ) How to make a good research presentation
- ( ) How to selecting research topics
- ( ) Computer science paper reading
- ( ) Research paper and thesis writing
- ( ) Computer science experiments
- ( ) Publishing your research work
- ( ) Job hunting tips and interview skills

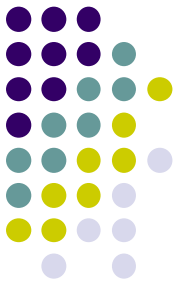
# Survey results



- ( 4.68 ) Research paper and thesis writing
- (4.21 ) Job hunting tips and interview skills
- (4.21 ) How to make a good research presentation
- (4.10 ) How to selecting research topics
- ( 4.05 ) Computer science experiments
- ( 4.00 ) Computer science paper reading
- ( 3.95 ) Publishing your research work
- ( 3.58 ) Introduction to computer scientists
- ( 2.79 ) Introduction to computer science

# Review

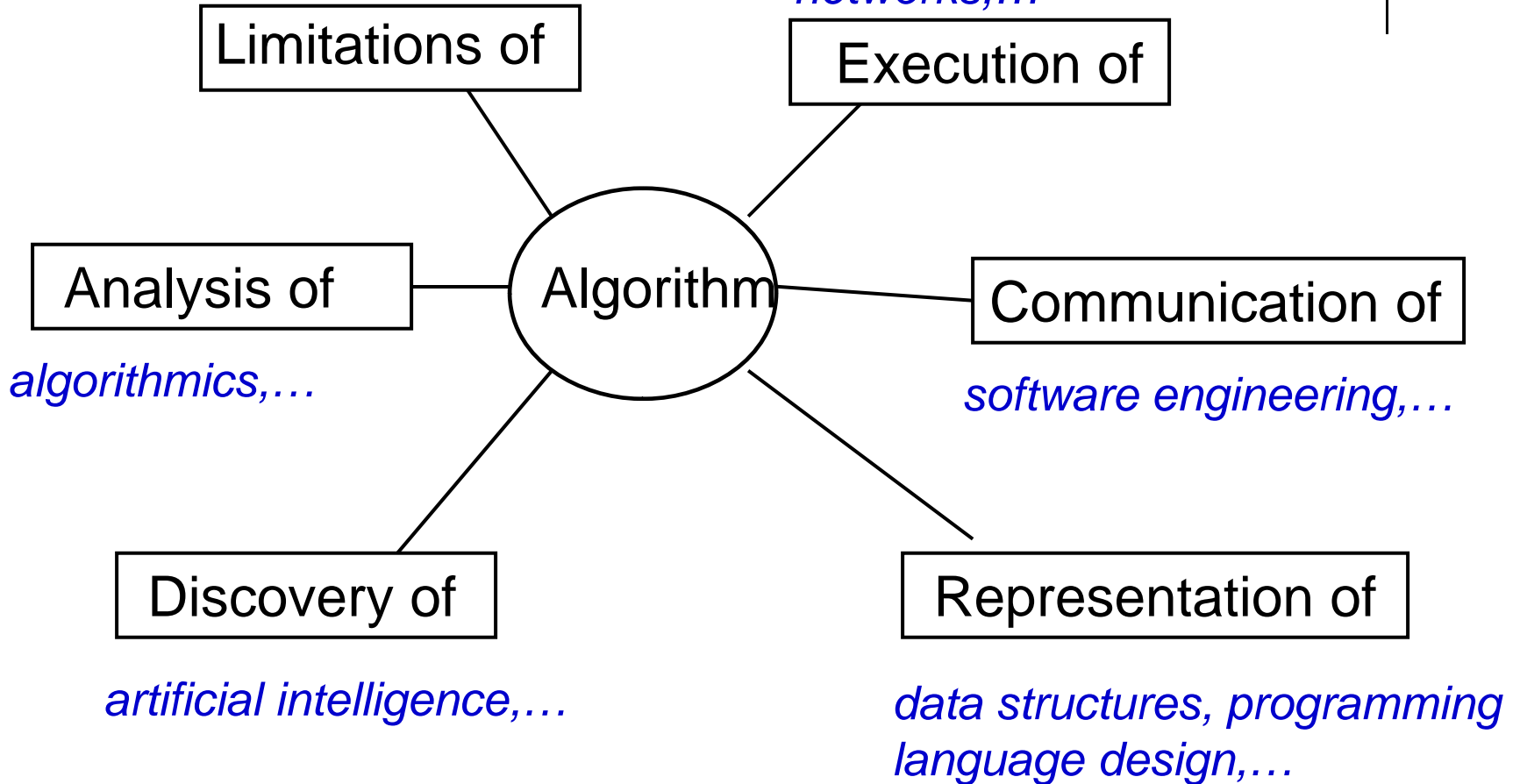
- What is Computer Science?
- Subfields of Computer Science



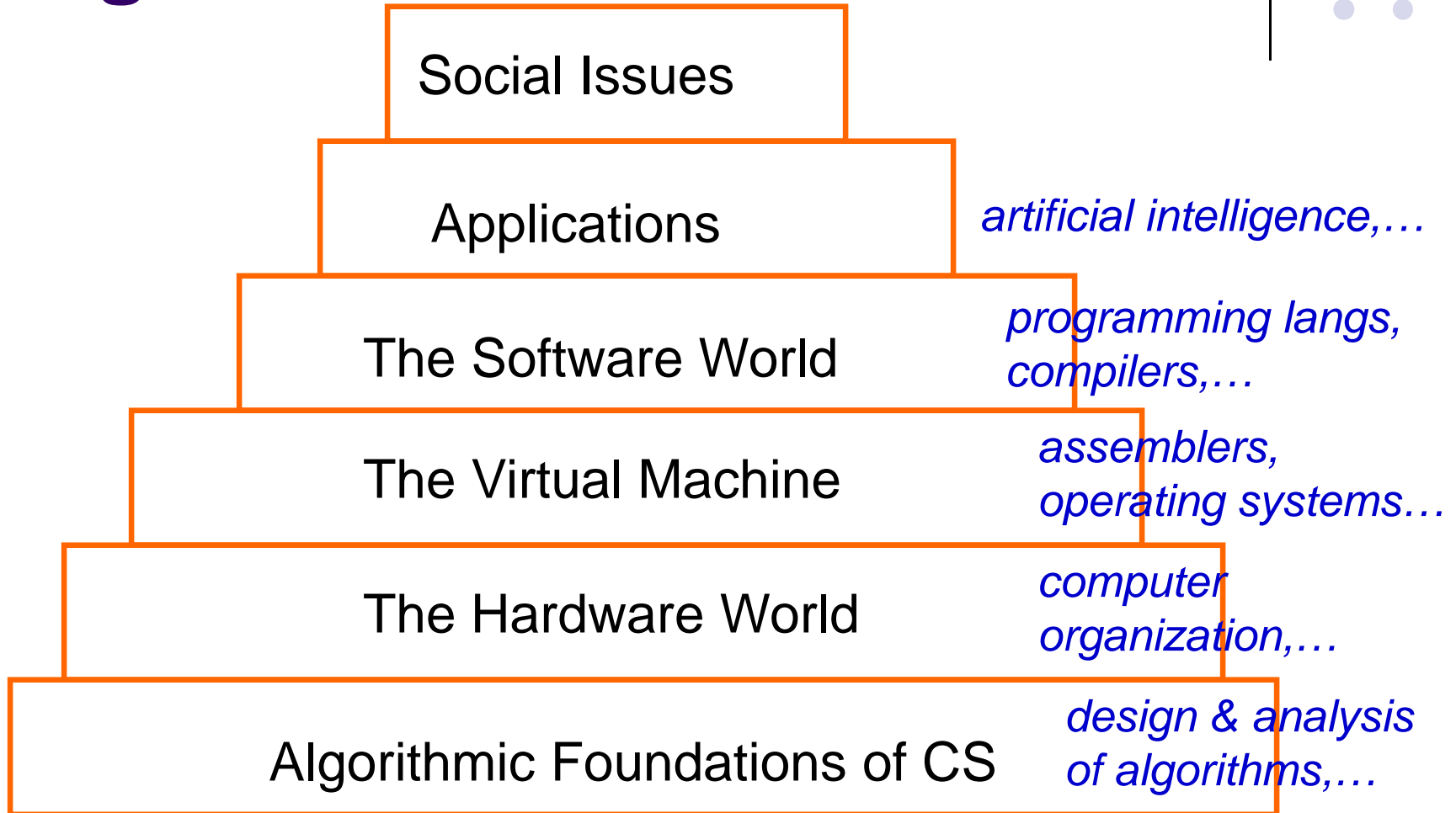
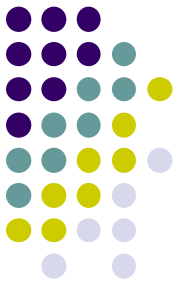
# Brookshear's Diagram

*theory of computation,...*

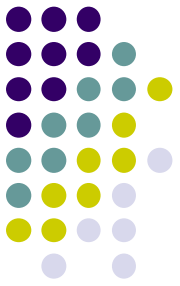
*architecture, operating systems,  
networks,...*



# Schneider & Gersting's Diagram



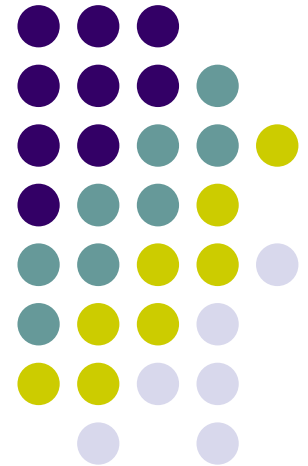
# Subfields of Computer Science

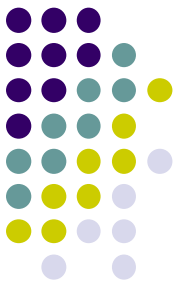


- Algorithms and Data Structures
- Architecture
- Operating Systems and Networks
- Software Engineering
- Artificial Intelligence and Robotics
- Bioinformatics
- Programming Languages
- Databases and Information Retrieval
- Graphics
- Human-Computer Interaction
- Computational Science
- Organizational Informatics

# Database Research

---

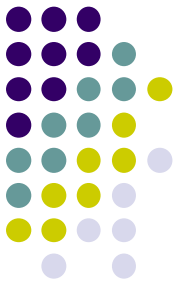




# Outline

- Five challenges on database research
  - Structured and unstructured data
  - Declarative programming
  - Database engine revisiting
  - Cloud data management
  - Mobile application
- Our research to meet those challenges

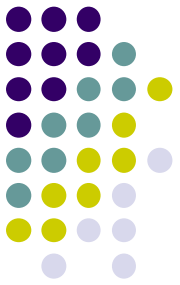
# 数据库的挑战:



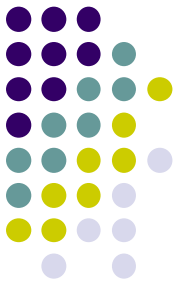
- Senior database researchers have gathered every few years to assess the state of database research and to recommend problems and problem areas deserve additional focus.
  - Laguna Beach, Calif. in 1989
  - Palo Alto, Calif. (“Lagunita”) in 1990 and 1995
  - Cambridge, Mass. in 1996
  - Asilomar, Calif. in 1998
  - Lowell, Mass . In 2003

# Claremont Meeting

- About 20 Database researchers
- Claremont Resort, Berkeley, CA  
May 29-30, 2008

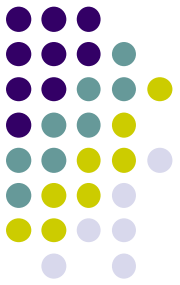


# The interplay of structured and unstructured data(1)

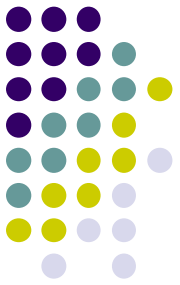


- Witnessing a growing amount of structured data
  - Millions of database hidden (Deep Web)
  - Millions of HTML tables and Mashups
  - Web 2.0 Service photo video websites

# The interplay of structured and unstructured data(2)



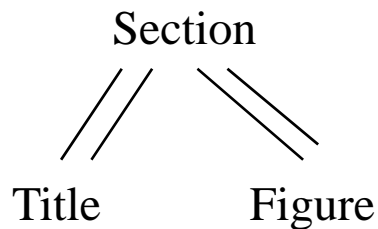
- Research challenge:
  - Extract structured meaning for unstructured data (IR, ML)
  - Querying and deriving insight from heterogeneous data
    - Keyword queries
    - Pay-as-you-go fashion



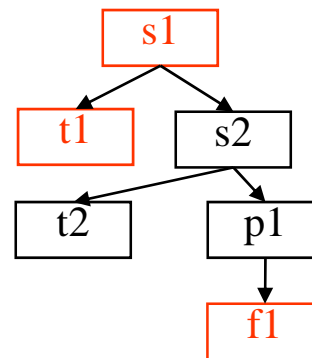
# XML search (1)

- XML twig query processing (SIGMOD'05, VLDB'05)
  - Problem Statement
    - Given an XML twig pattern  $Q$ , and an XML database  $D$ , we need to find **ALL** the matches of  $Q$  on  $D$ .

Twig pattern:



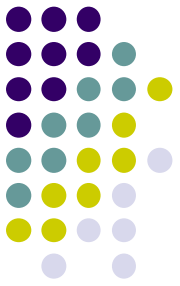
An XML tree:



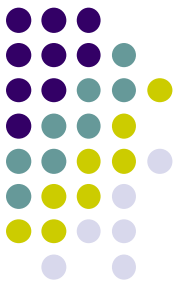
Query answers:

(s1, t1, f1)  
(s2, t2, f1)  
(s1, t2, f1)

# XML search (2)

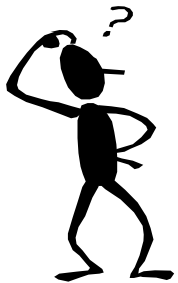


- XML keyword search (ICDE'09)
  - Problem Statement
    - How to efficiently rank the results of XML keyword query
  - Contribution:
    - Extend TF/IDF by incorporating the structure of XML data



# Approximate string search

- Approximate string queries (ICDE'08,09)
  - Problem Statement
    - Given a collection of string data, how to efficiently perform approximate search



Schwarzenger

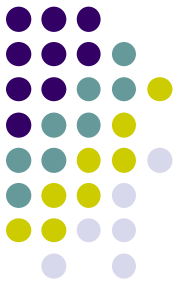
Search



Star
Keanu Reeves
Samuel Jackson
<b>Schwarzenger</b>
...

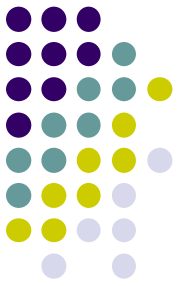
Output: strings  $s$  that satisfy  $\text{Sim}(q,s) \leq \delta$

# Revisiting database engines



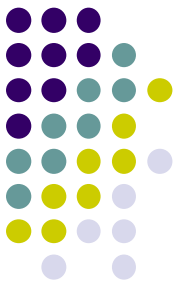
- Research topics
  - Remote RAM and flash as persistent media
  - Treat query optimization and physical data as a unified, adaptive, self-tuning task
  - Compressing and encrypting data with query optimization
  - Designing systems that embrace non-relational data models

# Declarative programming for Emerging platforms (1)

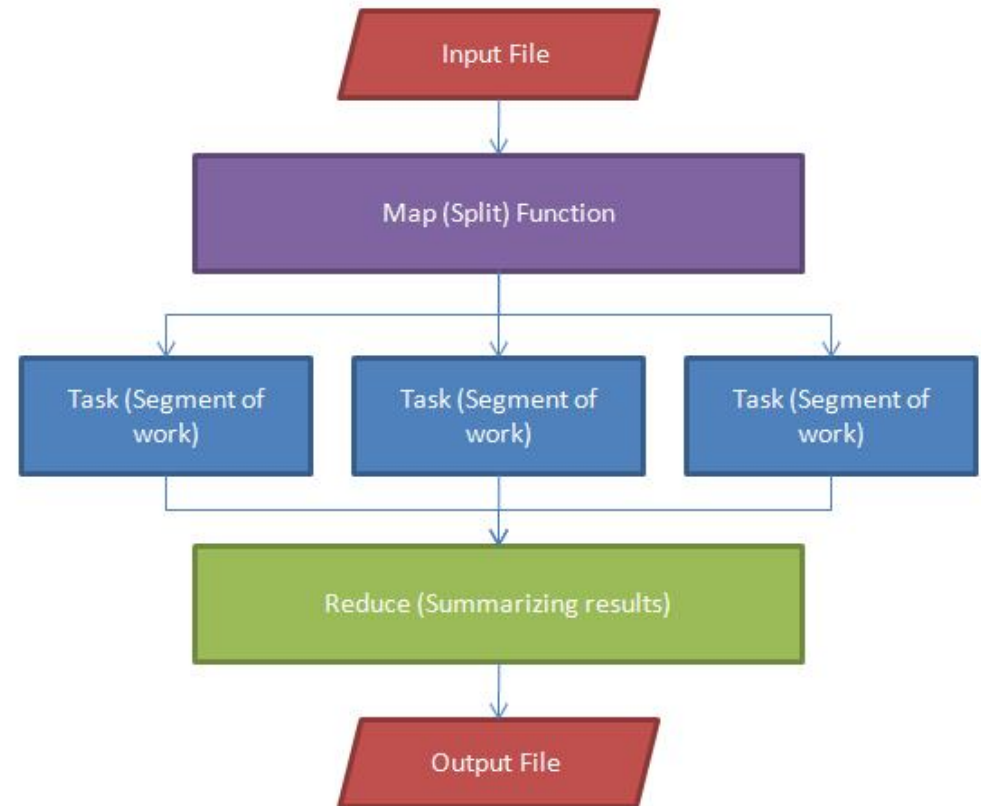


- Data-centric approach for emerging platforms
  - Manycore chips
  - Distributed services
  - Cloud computing platforms
  - .....

# Declarative programming for Emerging platforms (2)

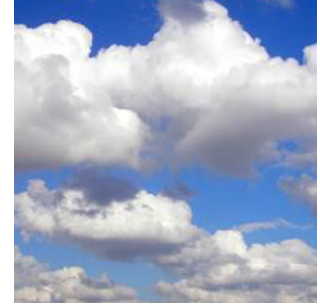


- Good examples
  - Map-reduce: data-parallelism
  - Ruby, Rails query-like logic
  - XQuery





# Cloud data management (1)

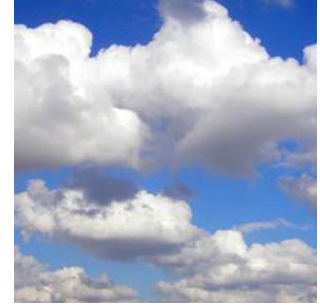


- Cloud service: shared commodity hardware for computing and storage
  - Application service (salesforce.com)
  - Storage service (Amazon Web service)
  - Computing service (Google App Engine)
  - Data service (Microsoft SQLServer data center)



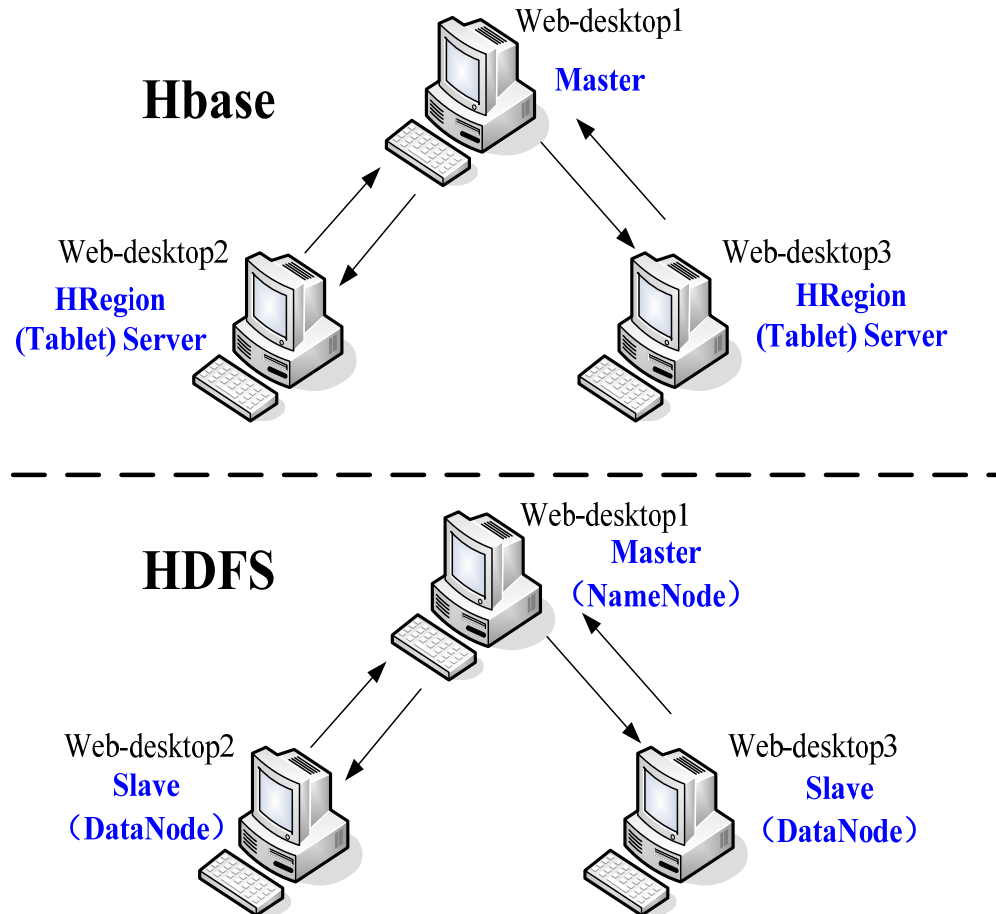
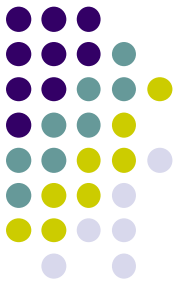


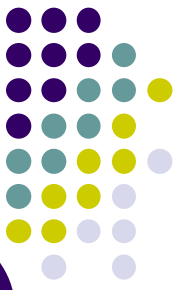
# Cloud data management (2)



- Research challenge
  - Self-management database: limited human invention, various workloads
  - Large scale query processing and optimization
  - Data security and privacy with sharing

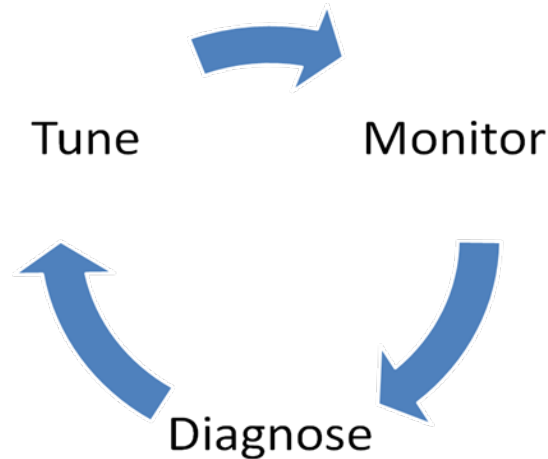
# Cloud data management



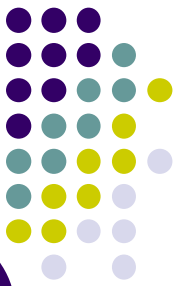


# Research topics about cloud data(1)

- Self management and self tuning



- Query optimization on thousands of nodes



## Research topics about cloud data(2)

- Source scheduling
  - Investigate the way the scheduling algorithm is currently implemented.
- Multi-Tenant-Efficient

# Mobile applications

- “On the go” interaction
- Location based service

