

<http://dmlab.xmu.edu.cn/post/bigdata>

*PPT*

# 第11章 大数据在互联网领域的应用

(PPT版本号: 2015年1月29日版本)

E-mail: [ziyulin@xmu.edu.cn](mailto:ziyulin@xmu.edu.cn) ▶▶

<http://www.cs.xmu.edu.cn/linziyu>





# 提纲

- 11.1 推荐系统概述
- 11.2 推荐算法 – 协同过滤
- 11.3 协同过滤实践 – 电影推荐系统

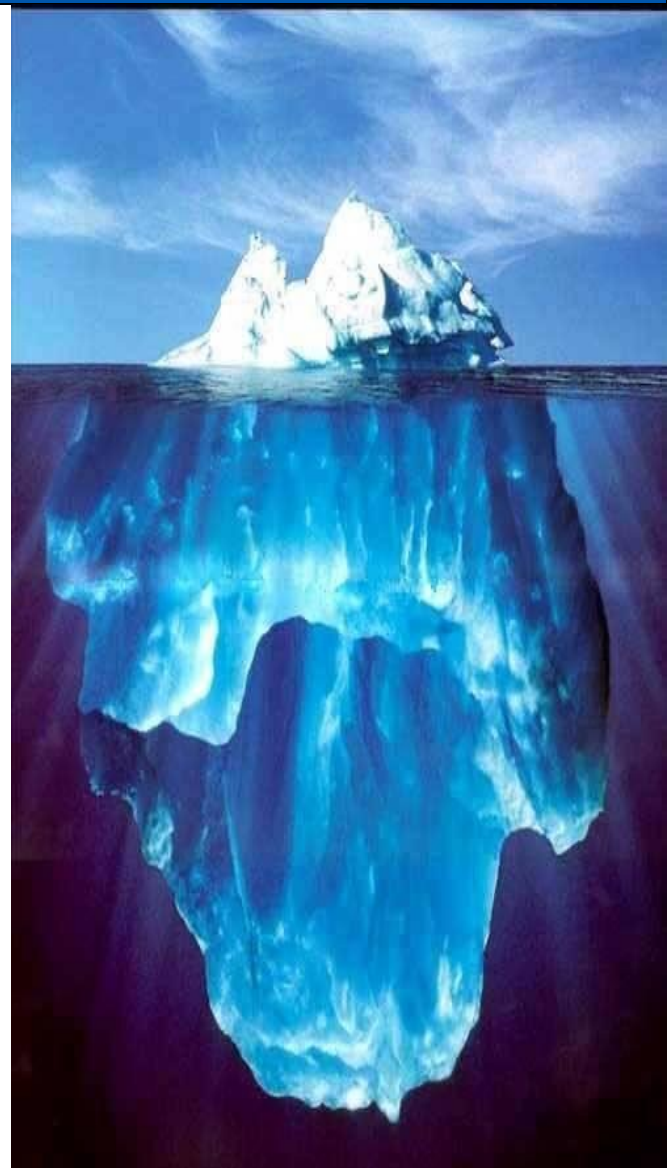
PPT  
21

---

2015 6 1

ISBN:978-7-115-39287-9

<http://dmlab.xmu.edu.cn/post/bigdata>





# 11.1 推荐系统概述

- 11.1.1
- 11.1.2
- 11.1.3
- 11.1.4
- 11.1.5



# 11.1.1 什么是推荐系统

- 

-



## 11.1.2 长尾理论

- 2004
- 
-



## 11.1.2 长尾理论

- 

-



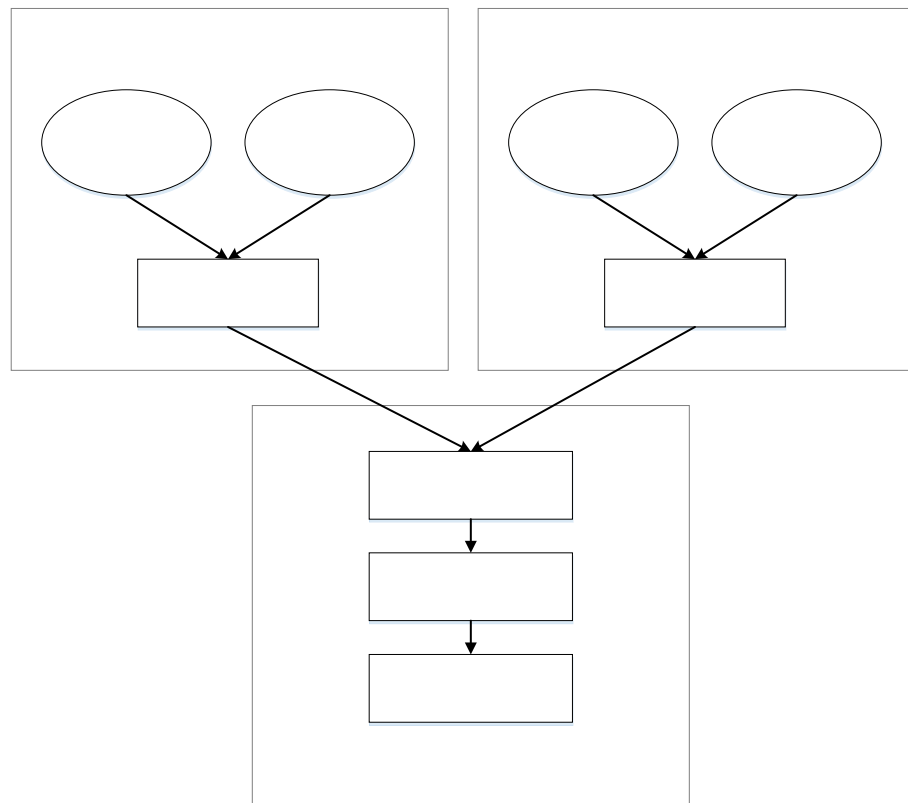
# 11.1.3 推荐方法

- - 
  - 
  - 
  - 
  - 
  -



# 11.1.4 推荐系统模型

3



11-1





# 11.1.5 推荐系统的应用

- 
- 

## 您最近查看的商品和相关推荐

根据您的浏览历史记录推荐商品

第 1 页, 共 10 页 [第一页](#)

				
				
beyerdynamic...	LEGO 乐高 创意百变组 西...	JVC 碳素薄膜头戴耳机 S400	Sennheiser...	Duc d'O 迪克多...
★★★★★ (134)	★★★★★ (98)	★★★★★ (70)	★★★★★ (209)	★★★★★ (27)
¥ 980.00	¥ 212.00	¥ 178.23	¥ 299.00	¥ 118.00

11-2



# 11.1.5 推荐系统的应用

- 
- 

猜你喜欢 [更多](#)



今日推荐歌单



基于歌曲 Space Oddi...

VH1 Storytellers  
David Bowie



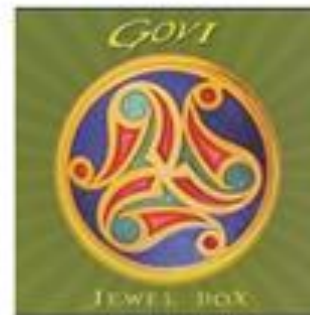
基于歌曲 Franz Liszt...

感動のショパン・ラ...  
李云迪



基于歌曲 Paganini: C...

The Master Violinist...  
Jascha Heifetz



基于歌曲 Romance ...

Jewel Box  
Govi



## 11.2 协同过滤

- 
- 
- 11.2.1 UserCF
- 11.2.2 ItemCF
- 11.2.3 UserCF ItemCF

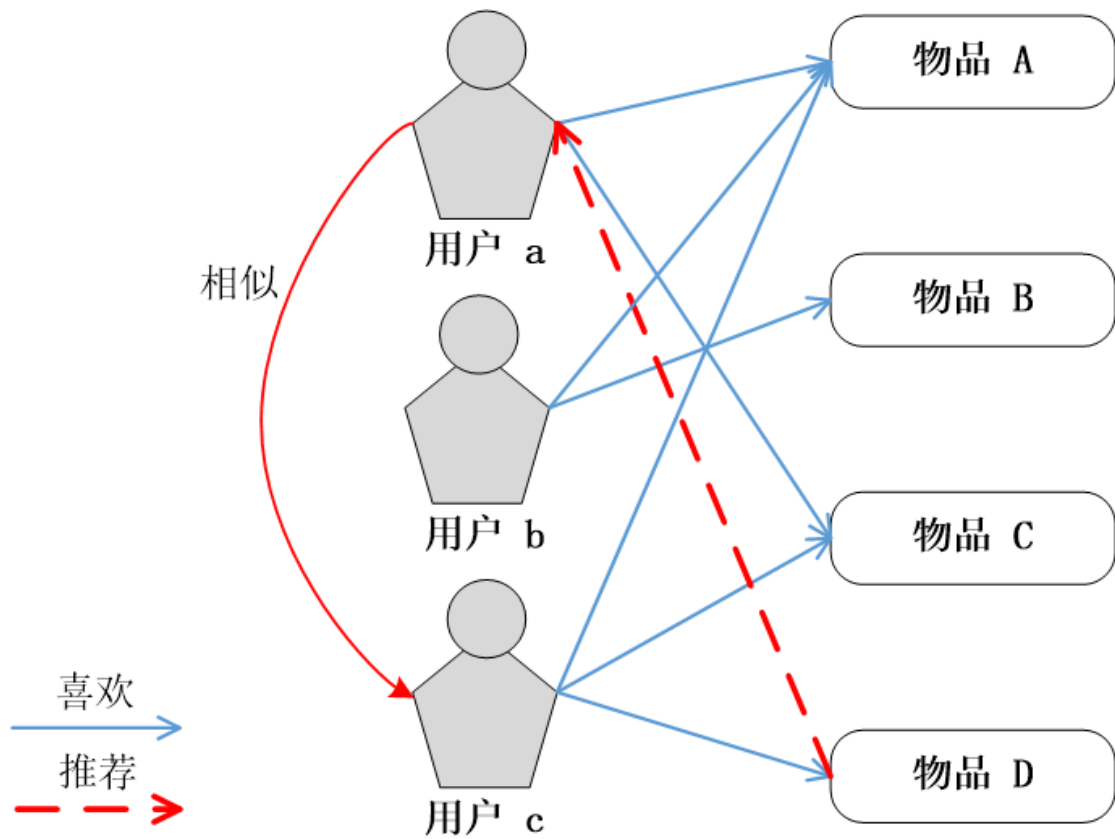


# 11.2.1 基于用户的协同过滤 (UserCF)

- UserCF 1992
- UserCF
- UserCF
  - 
  -



# 11.2.1 基于用户的协同过滤 (UserCF)



11-4

User CF



# 11.2.1 基于用户的协同过滤 (UserCF)

- UserCF

- Person Correlation Coefficient
- Cosine-based Similarity
- Adjusted Cosine Similarity

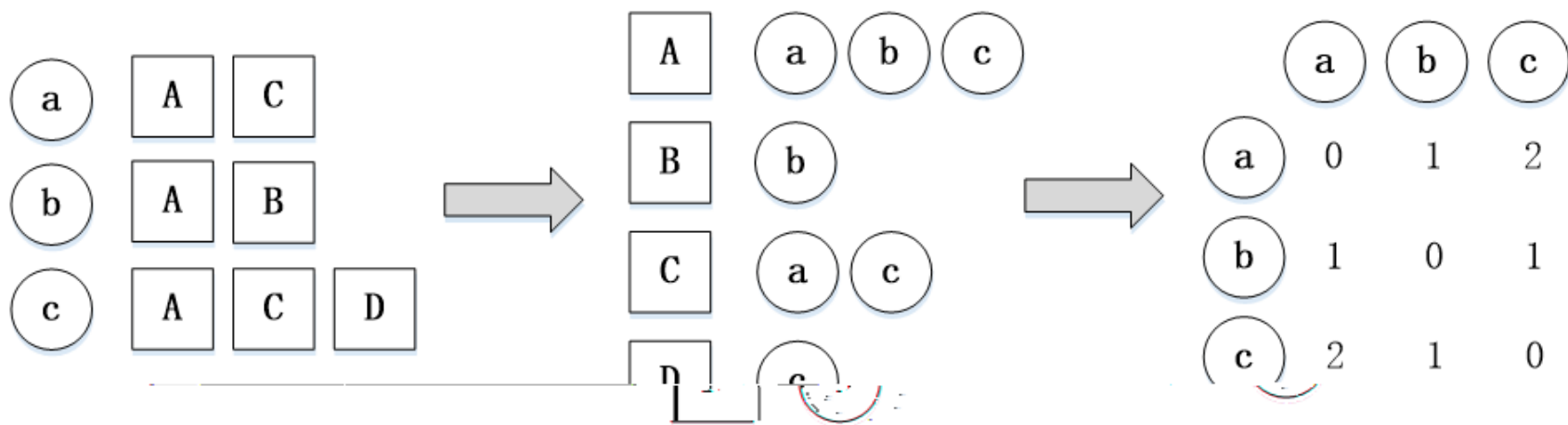
- $u$   $v$   $N(u)$   $u$   $N(v)$   
 $v$

$$w_{uv} = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)||N(v)|}}$$



# 11.2.1 基于用户的协同过滤 (UserCF)

- 
- 



(a) 用户喜欢的物品列表

(b) 物品对应的用户列表

(c) 相似度矩阵W



# 11.2.1 基于用户的协同过滤 (UserCF)

- $P_{ui}$   $u$   $i$
- $$p(u, i) = \sum_{v \in S(u, K) \cap N(i)} w_{uv} r_{vi}$$
- $S(u, K)$   $u$   $K$   $N(i)$   
 $v$   $i$   $w_{uv}$   $u$   $v$   $r_{vi}$   
 $r_{vi}=1$
- $P_{ui}$   $u$   $P_{ui}$   $N$   
 Top-N



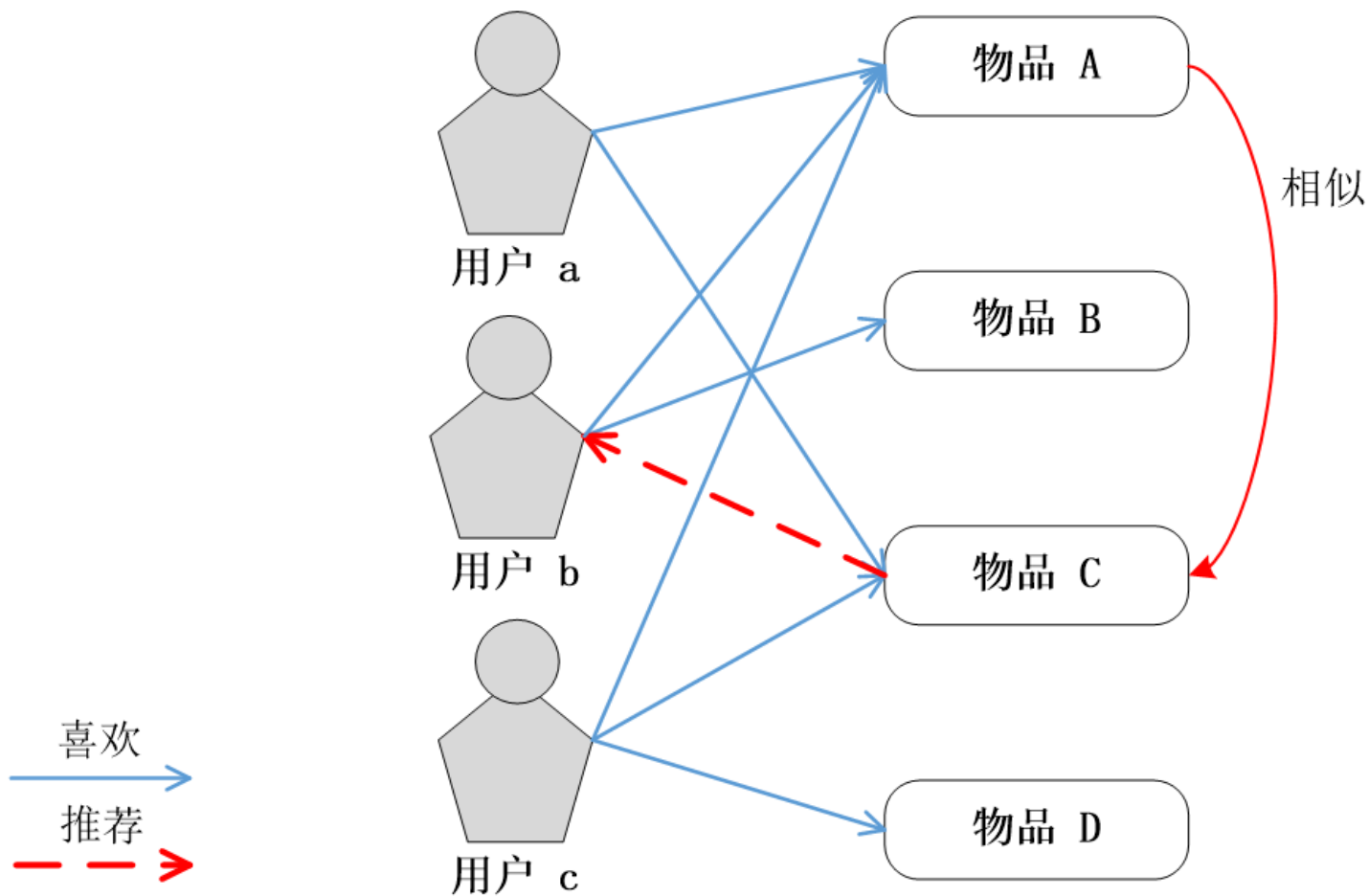


# 11.2.2 基于物品的协同过滤 (ItemCF)

- ItemCF  
Netflix ItemCF
- A A B B



# 11.2.2 基于物品的协同过滤 (ItemCF)



11-6

Item CF



## 11.2.2 基于物品的协同过滤 (ItemCF)

- ItemCF      UserCF

—  
—

- ItemCF

$P_{uj}$  ( UserCF )

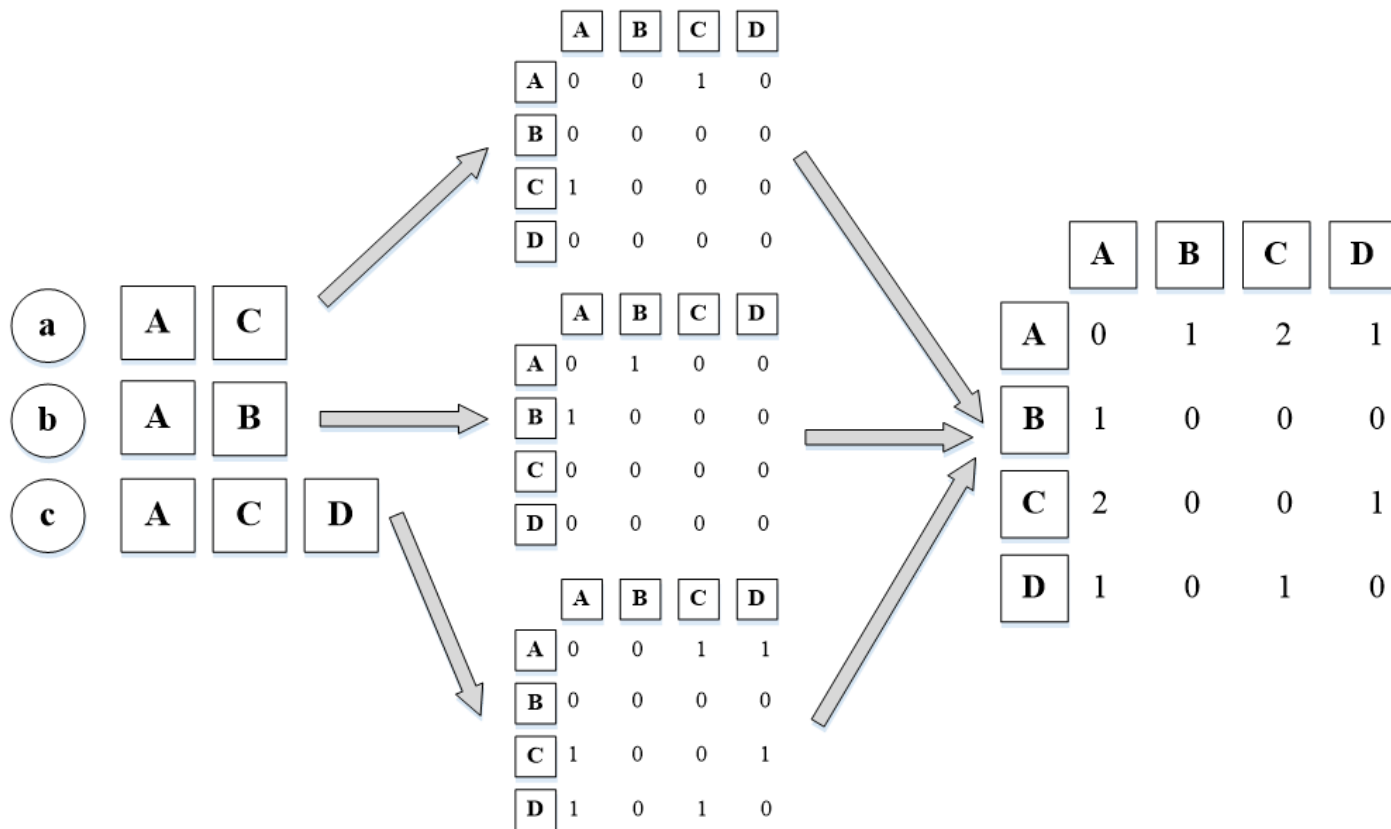
u      j

$$P_{uj} = \sum_{i \in N(u) \cap S(j, K)} w_{ji} r_{ui}$$



# 11.2.2 基于物品的协同过滤 (ItemCF)

- ItemCF



(a) 用户喜欢的物品列表

(b) 物品相似度矩阵M

(c) 物品相似度矩阵R



## 11.2.3 UserCF算法和ItemCF算法的对比

- UserCF                  ItemCF

- 

  - UserCF

  - ItemCF

- UserCF

- ItemCF



## 11.2.3 UserCF算法和ItemCF算法的对比

- UserCF
- UserCF  
UserCF
- ItemCF
- ItemCF



# 11.3 协同过滤实践

- 11.3.1
- 11.3.2
- 11.3.3
- 11.3.4
- 11.3.5



# 11.3.1 实践背景

- MovieLens Python ItemCF
- MovieLens 100k 100000 20 1000 1700
- 
-





## 11.3.2 实践数据

- 
- 

ID

ID

用户ID	电影ID	评分	评分时间戳
196	242	3	881250949
186	302	3	891717742
22	377	1	878887116
244	51	2	880606923
166	346	1	886397596
298	474	4	884182806
115	265	2	881171488
253	465	5	891628467
305	451	3	886324817
6	86	3	883603013
... ..			



## 11.3.3 实践流程

- - 1.
    - 2.
    - 3.
    - 4.
- Top-N



## 11.3.3 实践流程

- ID 1 10  
film: Titanic (1997), rating: 3.89483842578  
film: Air Force One (1997), rating: 3.86732319398  
film: L.A. Confidential (1997), rating: 3.8255074736  
film: Winter Guest, The (1997), rating: 3.77287384289  
film: Postman, The (1997), rating: 3.7726583083  
film: Wag the Dog (1997), rating: 3.76181477552  
film: Fire Down Below (1997), rating: 3.75018044393  
film: Anna Karenina (1997), rating: 3.74833946959  
film: Leaving Las Vegas (1995), rating: 3.74818188227  
film: Shadow Conspiracy (1997), rating: 3.7351481072

11-10



# 本章小结

- 

- 

- 

UserCF

ItemCF

UserCF

ItemCF

- 

Python



# 附录：主讲教师



主讲教师：林子雨

E-mail: [ziyulin@xmu.edu.cn](mailto:ziyulin@xmu.edu.cn)

<http://www.cs.xmu.edu.cn/linziyu>

<http://dblab.xmu.edu.cn>



1978

2013

40

2000



# 附录：大数据学习教材推荐



13

Hadoop  
HBase NoSQL  
MapReduce

HDFS

Hadoop HDFS HBase MapReduce

<http://dblab.xmu.edu.cn/post/bigdata>

21世纪高等教育计算机规划教材

COMPUTER

## 大数据技术原理与应用

—概念、存储、处理、分析与应用

Principles and Applications of Big Data Technology—Big Data  
Conception, Storage, Processing, Analysis and Application

林子雨 编著

- 搭建起通向“大数据知识空间”的桥梁和纽带
- 构建知识体系、阐明基本原理、引导初级实践、了解相关应用
- 为读者在大数据领域“深耕细作”奠定基础、指明方向

中国工信出版集团 人民邮电出版社  
POSTS & TELECOM PRESS



# 附录：中国高校大数据课程公共服务平台



## 中国高校大数据课程 公共服务平台

<http://dblab.xmu.edu.cn/post/bigdata-teaching-platform/>



3 FLASH

21世纪高等教育计算机规划教材



# 大数据技术原理与应用

## ——概念、存储、处理、分析与应用

Principles and Applications of Big Data Technology—Big Data  
Conception, Storage, Processing, Analysis and Application

林子雨 编著

- 搭建起通向“大数据知识空间”的桥梁和纽带
- 构建知识体系、阐明基本原理、引导初级实践、了解相关应用
- 为读者在大数据领域“深耕细作”奠定基础、指明方向



中国工信出版集团

人民邮电出版社  
POSTS & TELECOM PRESS

Department of Computer Science, Xiamen University, 2016