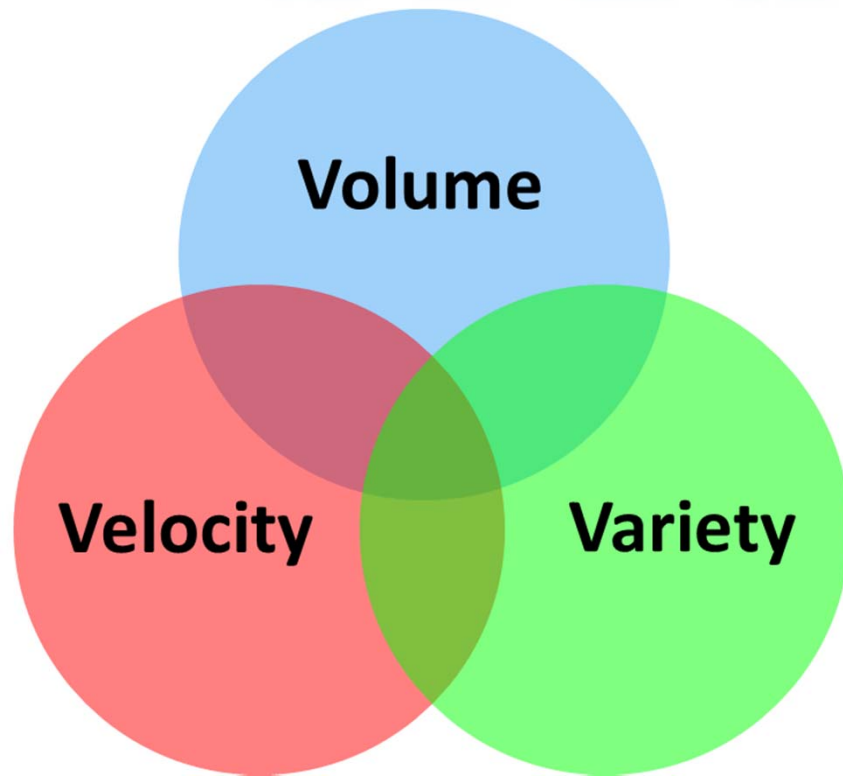


# 大数据系统 与大规模数据分析



陈世敏  
(中科院计算所)

孙翼  
(国科大计算机学院)



# 陈世敏

博士，中科院计算所研究员

- 分别于**1997**年和**1999**年获清华大学计算机系学士学位和硕士学位，于**2005**年获卡内基梅隆大学（**CMU**）计算机科学系博士学位；
- **2005-2013**年在美国英特尔研究院、卡内基梅隆大学和惠普研究院任研究员、高级研究员和科研经理；
- **2013**年加入中科院计算所
- 研究方向：数据库系统、大数据处理、体系结构

# 课程相关

- 课程时间地点

- 每周三晚（9-11节） 18:10 – 20:50

- 课间休息20分钟：时间调整到三节课的中间19:20-19:40

- 地点：教1-107

- 基础知识要求

- 数据库、程序设计、数据结构、计算机原理

- 考核方式：闭卷考试+作业

- 闭卷考试： 50%

- 作业1+作业2+作业3： 30%

- 大作业： 20%

- 课堂表现： +3%

# 课程安排

周次	内容
9次课	大数据系统 (陈世敏)
5次课	大规模数据分析 (孙翼)
第15周 (6月16日)	期末考试
第16周 (6月23日)	大作业验收, 课堂报告

# 课程安排（大数据系统部分）

周次	内容	
第1周	概论；关系型系统1：关系模型和关系运算	
第2周	关系型系统2：数据库的内部实现	
第3周	关系型系统3：事务处理和数据仓库	
第4周	大数据存储系统1：基础，文件系统，HDFS	作业1
第5周	大数据存储系统2：键值系统	
第6周	大数据存储系统3：图存储，document store	
第7周	大数据运算系统1：MapReduce, 图计算系统	作业2
第8周	大数据运算系统2：图计算系统, MapReduce+SQL	
第9周	大数据运算系统3：内存计算系统，Spark	大作业

# 课程安排（大规模数据分析部分）

周次	内容	
第10周	最邻近搜索和位置敏感（LHS）哈希算法	
第11周	数据空间的维度约化	
第12周	推荐系统	作业3
第13周	流数据采样与估计、流数据过滤与分析	
第14周	大规模数据建模平台应用举例：教育大数据的建模与分析	

2020—2021 学年春季和夏季学期校历



年度 year	2021 春季学期 (Spring Semester)																2021 夏季学期 (Summer Semester)			
月份 month	三月 (Mar)				四月 (Apr)				五月 (May)				六月 (Jun)				七月 (Jul)			
周次 week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	I (17)	II (18)	III (19)	IV (20)
星期一 (Mon)	8	15	22	29	5	12	19	26	3	10	17	24	31	7	14 端午 午节	21	28	5	12	19
星期二 (Tue)	9	16	23	30	6	13	20	27	4	11	18	25	1	8	15	22	29	6	13	20
星期三 (Wed)	10	17	24	31	7	14	21	28	5	12	19	26	2	9	16	23	30	7	14	21
星期四 (Thu)	11	18	25	1	8	15	22	29	6	13	20	27	3	10	17	24	1	8	15	22
星期五 (Fri)	12	19	26	2	9	16	23	30	7	14	21	28	4	11	18	25	2	9	16	23
星期六 (Sat)	13	20	27	3	10	17	24	1 劳 动节	8	15	22	29	5	12	19	26	3	10	17	24
星期日 (Sun)	14	21	28	4 清 明节	11	18	25	2	9	16	23	30	6	13	20	27	4	11	18	25
说 明	1、2021 年 3 月 8 日正式上课； 注：教职工放假由党政办统一安排。 2、2021 年 6 月 28 日—7 月 25 日为夏季学期。																			