

Hadoop与大数据： 一个数据库学者的解读

**Hadoop vs. Big Data:
Unscrambling of a Database Researcher**

周傲英



华东师范大学



提纲



- Hadoop的起源和历史
- Hadoop在数据管理中的地位
- 传统数据库发展回顾
- 什么是“大数据”
- 从数据库角度看大数据研究
- 结语

提纲

- Hadoop的起源和历史
- Hadoop在数据管理中的地位
- 传统数据库发展回顾
- 什么是“大数据”
- 从数据库角度看大数据研究
- 结语

Hadoop史前

- 2002: 开源搜索引擎Nutch
 - ▣ 2003: 索引1亿个网页
- 2003: GFS (Google 文件系统)论文发表
 - ▣ Sanjay Ghemawat, Howard Gobioff, Shun-Tak Leung:
The Google file system. SOSP 2003: 29-43
- 2004: MapReduce论文发表
 - ▣ Jeffrey Dean, Sanjay Ghemawat: MapReduce:
Simplified Data Processing on Large Clusters. OSDI
2004: 137-150

Hadoop诞生

- 2004 – 2006: Nutch + DFS + MapReduce
 - ▣ Doug Cutting and Michael J. Cafarella
- 2006.1 – 2008: **Web-scale Hadoop!** (@Yahoo!)

Hadoop生态圈（部分）

- 2006: BigTable论文发表
 - ▣ Fay Chang, Jeffrey Dean, Sanjay Ghemawat, et al:
Bigtable: A Distributed Storage System for Structured Data. OSDI 2006: 205-218
- 2006 – 2008: HBase开发开始到成为Hadoop子项目
- 2008.9: HIVE成为Hadoop子项目
- 2006: Pig @Yahoo!

Hadoop生态圈（部分）

- 2006: BigTable论文发表
 - ▣ Fay Chang, Jeffrey Dean, Sanjay Ghemawat, et al:
Bigtable: A Distributed Storage System for Structured Data. OSDI 2006: 205-218
- 2006 – 2008: HBase开发开始到成为Hadoop子项目
- 2008.9: HIVE成为Hadoop子项目
- 2006: Pig @Yahoo!

提纲

- Hadoop的起源和历史
- Hadoop在数据管理中的地位
- 传统数据库发展回顾
- 什么是“大数据”
- 从数据库角度看大数据研究
- 结语

Hadoop在数据管理中的地位

- Hadoop具有新颖的技术思路
- 在处理网页数据等方面取得巨大成功
- Hadoop经过几年的发展，从一门边缘技术成长为一种事实上的标准。

“The Great MapReduce Debate”

- “MapReduce: A Major Step Back”, Database Column Blog, Jan. 17, 2008



- ▣ A giant step backward in the **programming paradigm** for large-scale data intensive applications
- ▣ A sub-optimal **implementation**, in that it uses brute force instead of indexing
- ▣ Not novel at all -- it represents a specific implementation of **well known techniques** developed nearly 25 years ago
- ▣ **Missing** most of the **features** that are routinely included in current DBMS
- ▣ **Incompatible** with all of the tools DBMS users have come to depend on

“The Great MapReduce Debate”

- “MapReduce: A Major Step Back”, Database Column Blog, Jan. 17, 2008



- 主要针对**MapReduce**的开源实现——**Hadoop**
- A giant step backward in the **programming paradigm** for large-scale data intensive applications
- A sub-optimal **implementation**, in that it uses brute force instead of indexing
- Not novel at all -- it represents a specific implementation of **well known techniques** developed nearly 25 years ago
- **Missing** most of the **features** that are routinely included in current DBMS
- **Incompatible** with all of the tools DBMS users have come to depend on



Be lover, no fighter!



提纲



- Hadoop的起源和历史
- Hadoop在数据管理中的地位
- 传统数据库发展回顾
- 什么是“大数据”
- 从数据库角度看大数据研究
- 结语

什么是数据库?

□ Data+Base

- Oracle/DB2/MySQL
- 1974: System R
- 1968: IMS (IBM Information Management System)
 - 层次模型
- 1965: IDS (Integrated Data Store)
 - 网状模型
- 1965: List Processing Task Force under CODASYL
 - 1967: LPTF → DBTG (Data Base Task Group)
- 1959: CODASYL (Conference on Data Systems Languages)

数据库的基本目的

- 计算机从数值计算转到数据处理，OS中的文件系统发展成DBMS
 - COBOL, CODASYL
 - **DBTG报告** (图灵奖获得者Charles Bachman)
 - 数据库系统最原始的目标是解决**记账(Billing)**和**订票(Booking)**问题
 - 商务智能的雏形
 - 事务处理 (图灵奖获得者Jim Gray)
- 数据库发展伊始，其目标应用简单明确**

数据库的三大成就

- 关系模型
 - ▣ E.F. Codd (数据库领域第二个图灵奖获得者)
- 事务处理
 - ▣ Jim Gray (数据库领域第三个图灵奖获得者)
- 查询优化
 - ▣ 结构化 (模式和实例分离)
 - ▣ 关系数据库理论 (关系代数)
 - ▣ 物理存储 (索引/统计)

成功造就了数百亿美元的数据库产业

数据库研究的拓展（挣扎）

- 数据建模
- 扩展的数据库系统 (EDBT)
- OLAP
- 数据仓库
- 数据挖掘

数据库研究的拓展（挣扎）

□ 数据模型（查询语言及查询优化）

- 非一范式模型 (Non First Normal Form, NF²,)
- 集合值数据模型 (Set-Valued Data Model)
- 复杂对象数据模型 (Complex Object Model)
- 泛关系模型 (Universal Relational Data Model)
- 图数据模型 (Graph Data Model)
- 面向对象数据模型/演绎数据库 (OODB/DeDB)
- XML/OEM (Object Exchange Model)

试图突破RDB的局限，解决BOM等问题

数据库研究的拓展（挣扎）

性能

- 实时数据库
- 内存数据库
- 数据库机
-

试图突破RDBMS
的性能/功能局限,扩
大应用范围

功能

- 数据库程序设计语言 (DBPL)
- 科学统计数据库 (SSDB)
- 演绎数据库 (While子句, DBLP)
- 对象关系数据库 (BLOB)
- 面向对象数据库
- 多媒体数据库
- 工程数据库

似曾相识?

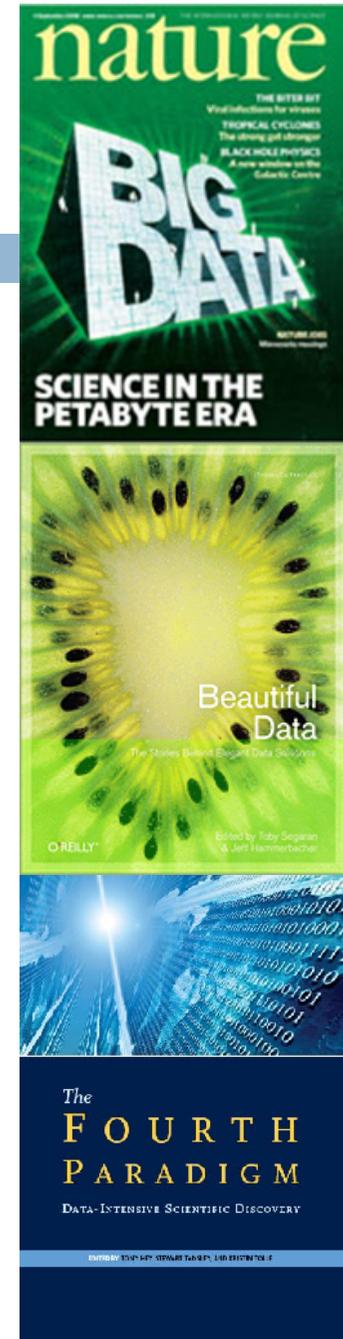
- MapReduce vs. Pig Latin vs. HBase/HIVE的类SQL语法
 - 网状模型 vs. 层次模型 vs. SQL
 - 模型与语言
- HBase/HIVE/Cassandra
 - System R/INGRES
 - 类似接口、多种系统
- Spark/Storm
 - 实时数据库/数据流处理
 - 实时、高效实现
- Mahout
 - OLAP/数据仓库
 - One-Size-Fit-ALL

提纲

- Hadoop的起源和历史
- Hadoop在数据管理中的地位
- 传统数据库发展回顾
- 什么是“大数据”
- 从数据库角度看大数据研究
- 结语

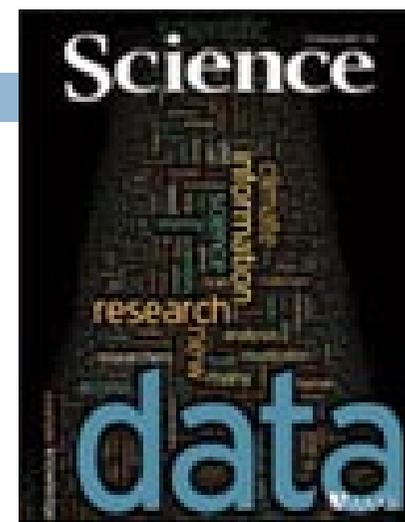
“大数据”概念的诞生

- 2008年9月4日《自然》(Nature)刊登了一个名为“**Big Data**”的专辑
- 2009年7月O'Reilly Media出版社出版了一本名为“**Beautiful Data**”的书
- 2009年10月微软为纪念Jim Gray, 出版了“**第四范式—数据密集的科学发现**(**The Fourth Paradigm—Data Intensive Scientific Discovery**)”



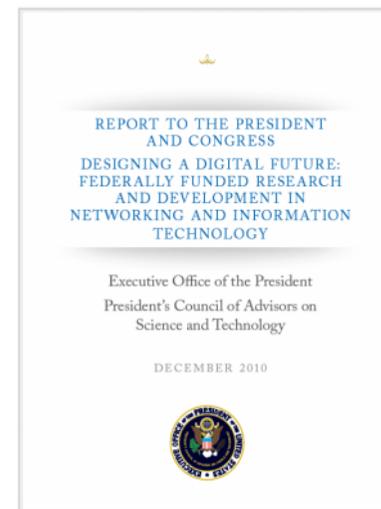
“大数据”概念的诞生

- 2011年2月11日：Science刊登了一个名为**Dealing with Data**的专辑，联合Science: Signaling、Science: Translational Medicine和Science Careers推出相关专题，讨论数据对科学研究的重要性
- 同一天，在美国很受欢迎的智力竞答“危险边缘（**Jeopardy**）”电视节目中，IBM的“**沃森**”系统以绝对优势战胜两名人类顶级选手。
- 和14年前的“深蓝”（战胜加里·卡斯帕罗夫）相比，“沃森”除具有超群的计算能力外，更拥有超大规模的数据以及数据处理能力



“大数据”概念的诞生

- 2012年3月29日，美国总统科技政策办公室OSTP（Office of Science and Technology Policy）宣布了每年投资两亿美元的“大数据研究计划”（Big Data R&D Initiative）
- 同一天，我国科技部发布的“‘十二五’国家科技计划信息技术领域2013年度备选项目征集指南”把大数据研究列在首位



“大数据”概念的诞生

- “大数据”主要讨论的是与**科学发现**有关的数据
- “大数据”很好地概括了当前数据管理领域问题的重要性和多样性
- 数据问题在研究和应用上空前的深度和广度
- “大数据”成为一个时髦的术语(Buzzword), 成为工业界与学术界共同关注的热点
- “大数据”是个合适的**umbrella**, 较广的覆盖面
- **应用驱动**, 将IT的重点转移到数据方面, 超越传统数据库的理念

“大数据” 概念的内涵

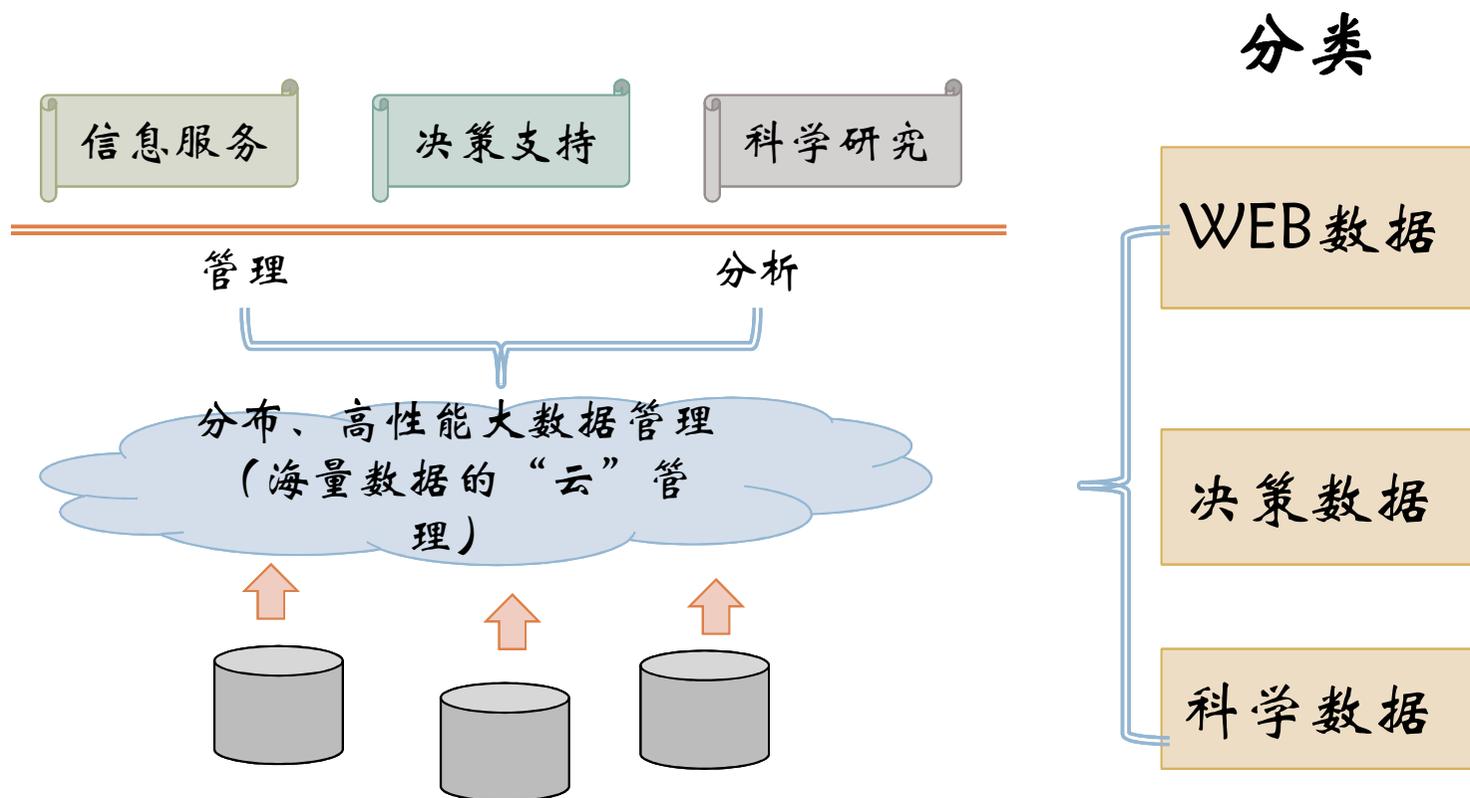
- 四个“V”
 - Volume
 - Variety
 - Velocity
 - Veracity

“大数据”概念的外延

- 与应用密切相关的各类数据，强调对于支持实际应用所涉及到的多个来源且相互关联的大量、高速、异构、质差的数据
 - ▣ 生产数据、设计数据、统计数据
 - ▣ 文本、多媒体数据、各种文档数据
- Heterogeneous Information Network
- 世界上的数据80%是非结构化数据
- 80-20规则

大数据的分类

27

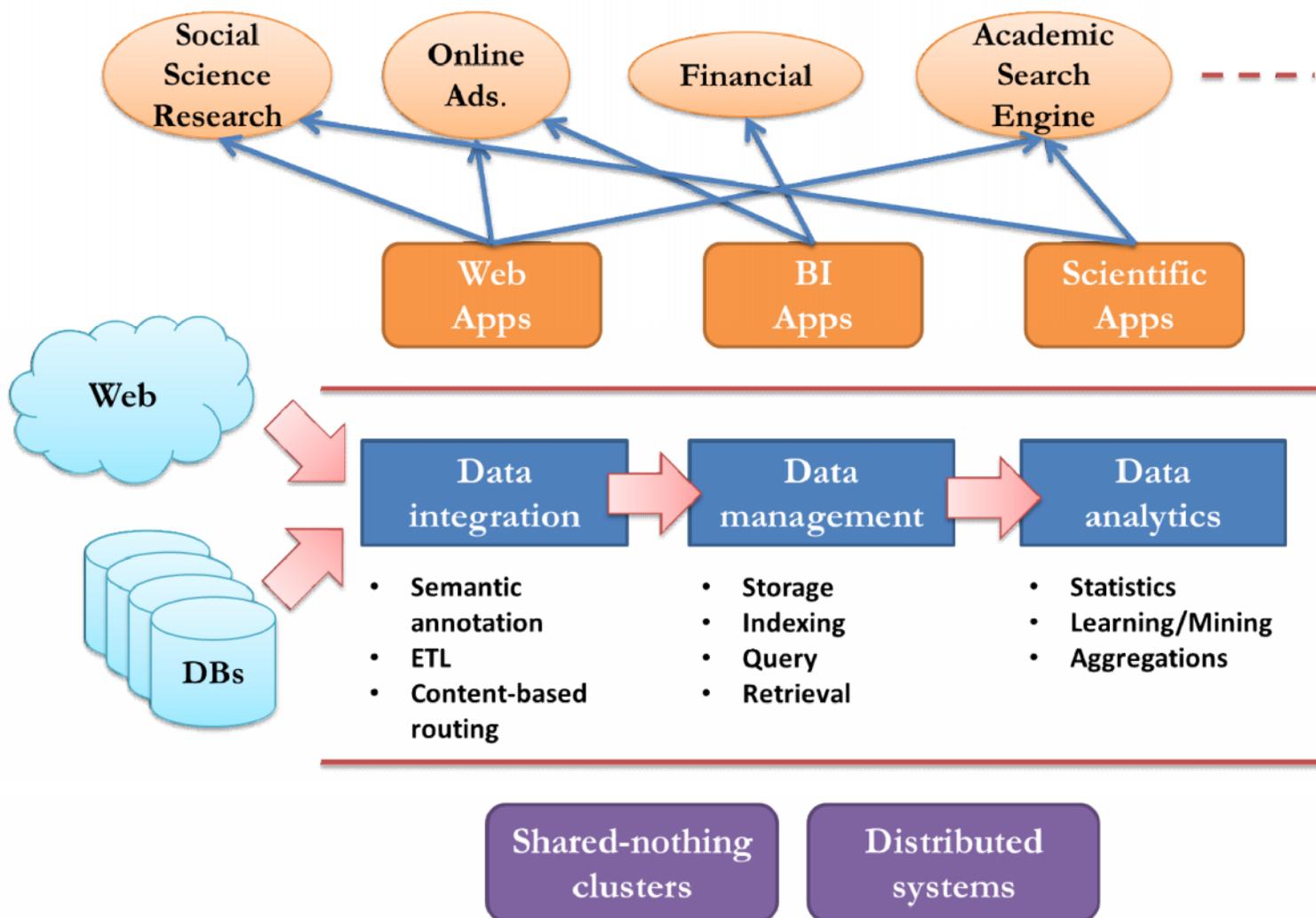


大数据的分类

28

- WEB数据
- 决策数据
- 科学数据
 - 科学实验
 - 科学观测
 - 科学文献
 - 设计数据

大数据的应用



大数据 ?= Hadoop

- 来源于Google，在类似搜索引擎的查询并行化分析处理领域取得极大成功
- 针对大规模数据密集型应用的编程范式（programming Paradigm）
- 所基于的BigTable和HDFS是非常质朴的数据模型和存储系统
- 适用领域有限，为大数据研究打开了思路，但绝不代表大数据技术全部(仅是一个case)

回到起点来重新审视数据管理之目的

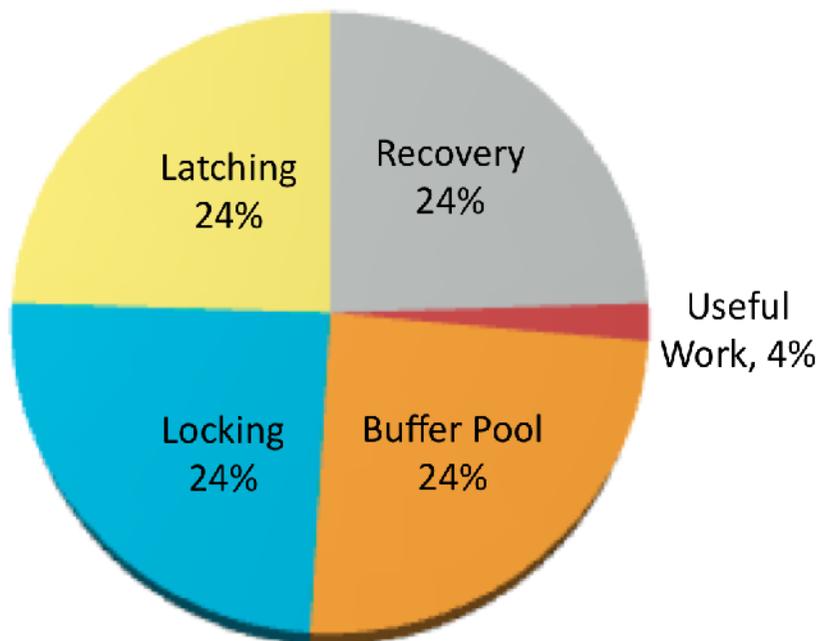
提纲

- Hadoop的起源和历史
- Hadoop在数据管理中的地位
- 传统数据库发展回顾
- 什么是“大数据”
- 从数据库角度看大数据研究
- 结语

传统数据库管理系统的问题

□ 大量系统开销(Shore DBMS上的实验)

- 各种锁和恢复机制
- 用于解决内外存的一致性问
题



数据管理的三个维度

- 数据
 - 结构化:
 - 半 / 非结构化
 - 无结构
- 管理和处理需求
 - 查询 / 检索 / 统计 / 挖掘
 - 离线 / 在线 (human real-time)
 - 即时 (on-demand/ad-hoc) / 连续
 - 本地 / 远程 / “云”?
- 硬件支撑
 - PC / 服务器 / 集群 / 分布式多集群
 - GPU / SSD / PCM / ...

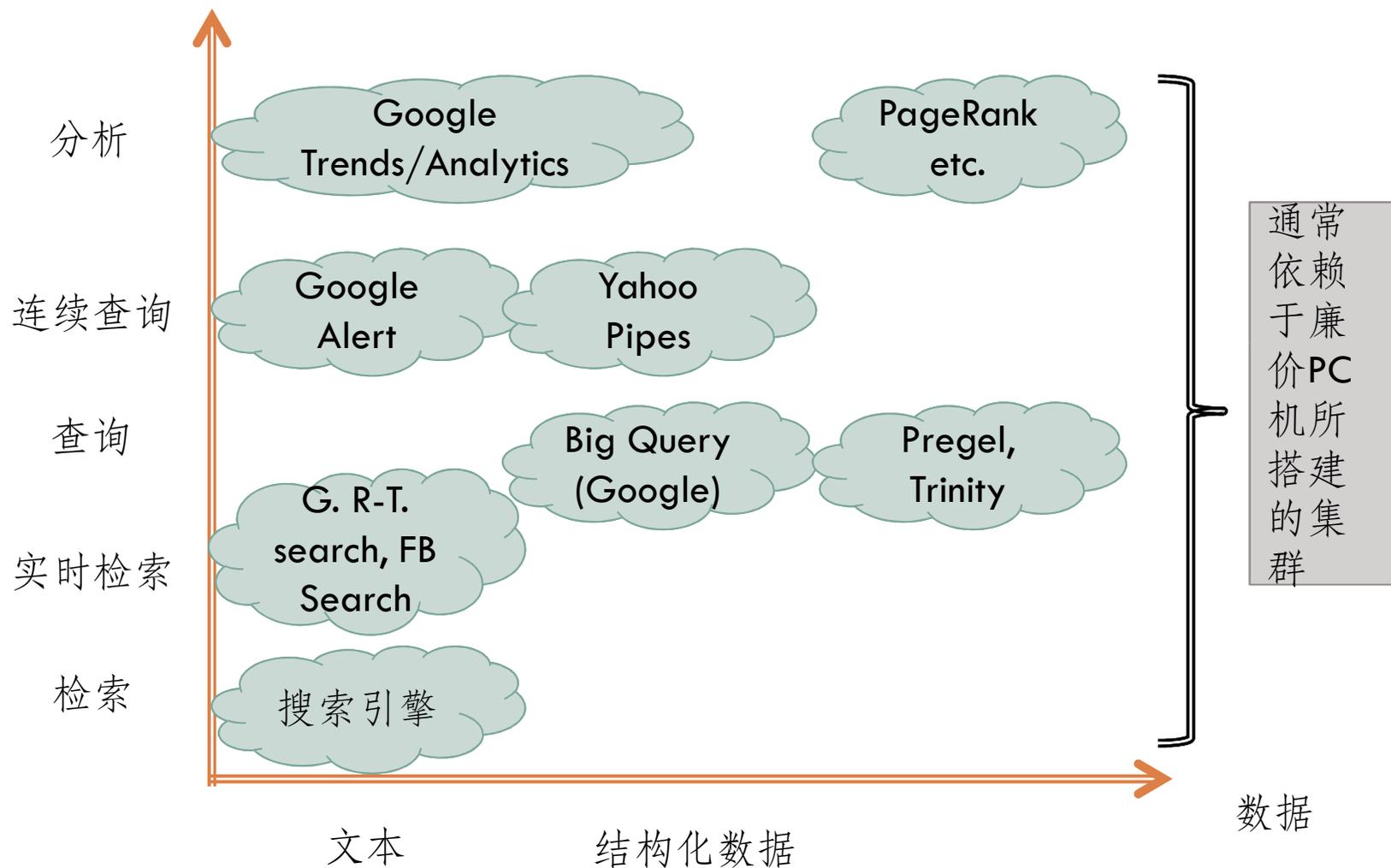
大数据和数据库的关系

34

- 数据库界从一开始就探索过，但还是过于保守
 - ▣ 忘不掉ACID，舍不得Relation，忽视实际应用
- 沉浸在自己的世界里
 - ▣ 空值理论(Null Value)，泛关系 (Universal Relation)
 - ▣ 数据库设计的范式理论 (FD, MVD, 4NF, 5NF,
 - ▣ 潜意识地奉行“一招鲜” (One Size Fits All, OSFA)
- Hadoop+Map/Reduce+Bigtable+HDFS响亮一击
- 回到起点来考虑数据管理问题，豁然开朗

大数据是数据库的自然延伸

Web数据管理



Web数据管理

- 新一代搜索引擎
- 云计算和MPP架构
- Hadoop+Map/Reduce
- Bigtable+HDFS
- Web数据应用
 - ▣ 在线广告, 社交媒体
- 研究问题
 - ▣ 图模型, 高速算法

Web数据管理示例：社交媒体数据管理

□ 每次用户刷新页面，后台系统需要做什么？

1. 查询用户关注/订阅/喜爱的其它用户/应用

- 记录于“关注列表”

- 目标：如果用户x关注了用户y，那么x应该能够看到y所发布的消息

2. 根据窗口获取所需用户/应用行为和记录

- 记录于“活动列表”，存于缓存，更新快速、频繁

- 目标：获取x所关注的所有的y所发布的最新的消息

3. 拼接生成时间线 (timeline)

- 查询结果，须合并排序

- 目标：只获取需要在x页面上显示的部分，按时间排序

Web数据管理示例：社交媒体数据管理

wngjian
上海, 黄浦区

740 | 642 | 1208
关注 | 粉丝 | 微博

看似简单的页面显示背后的大量数据处理支持

有 32 条新微博，点击查看

陈染ChanIm: 回复@仓鼠的小玮巴:杜甫不是说了嘛:“同学少年多不见”
//@仓鼠的小玮巴:切, 不想见我就算了 //@陈染ChanIm:那就再也不见 //@仓鼠的小玮巴:毕业了, 让我们学着张小娴的语气, 永不说再见 🤔

@Jason林骏超是五五的林先生 : 兄弟们, 毕业了~~~@深大爆料 @深大微波炉 @深大小站

今天19:02 来自新浪微博

转发(80) | 评论(3)

10分钟前 来自新浪微博

转发 | 收藏 | 评论

推送

我的首页

@提到我的
关注列表
我的评论
join
活动列表

我的收藏

活动列表
join
活动列表

实时统计

节日

获取个性化内容

展示“消息链”

丰富信息量

sky1ine

BrooklynDecker

加关注 国际超模布鲁

换一换

Web数据管理示例：社交媒体数据管理

□ 难点

- T: 110M条消息/天 \approx 1200条/秒
19GB/天 (文本)
- W: 峰值32312条/秒
- W: 平均每个人关注540人
T: 平均每个人关注36人
- W: 转发超过1000的微博中，超过80%的转发发生在1个小时之内的占59%

要求：（准）实时，大规模并发
计算：查询，连接，统计

推送

关注列表
join
活动列表

活动列表
join
活动列表

实时统计

T: 

W:  新浪微博

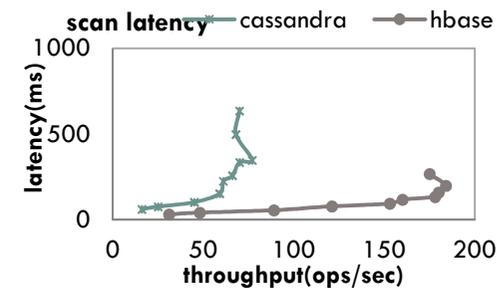
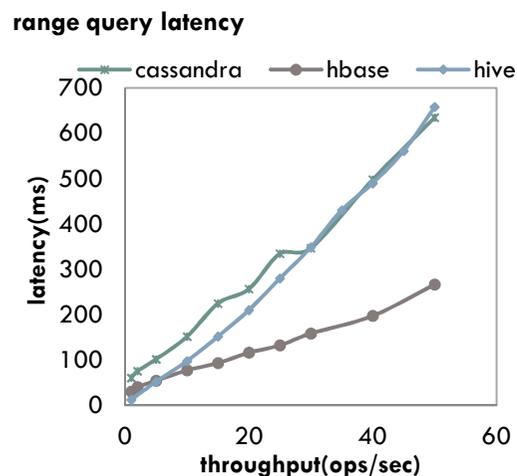
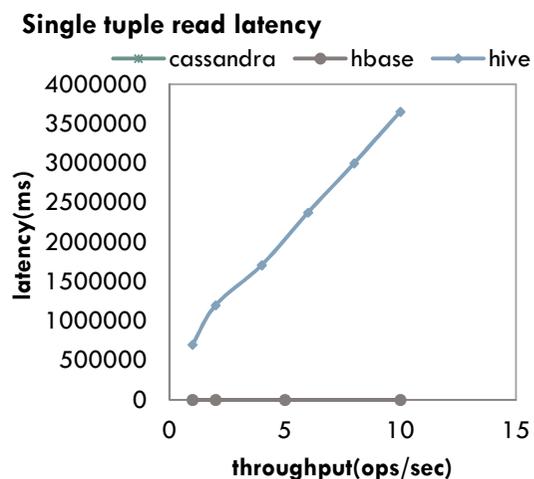
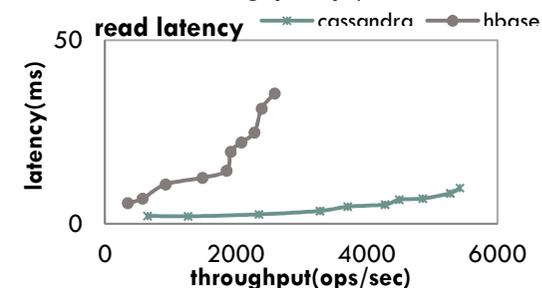
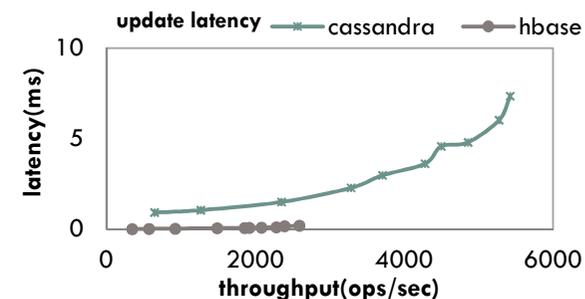
Web数据管理示例：社交媒体数据管理

问题：如何为每个用户的每次访问生成一个个性化的展示界面，对其所订阅的Timeline进行处理，展示分类/聚类/摘要后的内容？

- 现实应用，服务于：移动终端、舆情、营销
- 异常困难：
 - 先查询，再挖掘
 - 在快速变化的数据集上进行实时处理
 - 难以预处理
 - 复杂的分析处理
 - 内容挖掘/图结构挖掘

HBase/Cassandra/HIVE实验

- Web 2.0数据（社交媒体数据）需要均衡的高读写性能
- 现有系统优化目标单一
 - ▣ 优化读操作（HDFS,Cassandram,Hive）
 - ▣ 优化写操作（Hbase, Dynamo）



决策数据管理



决策数据管理

- 传统DBMS秉承的 one-size-fits-all 的理念不合适
- OLAP和数据仓库技术在新的硬件和体系结构情形下有新的发展机遇
 - ▣ 高可靠的MPP架构，提升系统的横向可扩展性
 - ▣ 结合内存发展趋势（NUMA架构，多层次Cache，共享Cache），实现内存数据的统一按列存储与管理（跨节点、跨CPU）
 - ▣ 结合CPU发展趋势（众核，超线程，内嵌GPU），采用新型分析执行模型（function at a core），优化分析并行执行度（分析内、分析间优化）
- 应用：实时商务智能
- 研究问题
 - ▣ 内存数据库，高性能集群，优化分析

决策数据管理示例: BI

- 传统处理方式
 - 离线: ETL, 物化视图, 报表生成, 规则提取 / 模型训练
 - 在线: OLAP
- 新问题: 实时即席决策分析
 - 数据来源于各种传感器
 - GPS, RFID, ...
 - 即席分析, 即时处理, 即时响应
 - 离线ETL不能满足需要
 - 应用: 智能电网、供应链管理、物流优化、...

决策数据管理示例: SSE POC

- 三个月交易数据 (850GB左右)
- 硬件 IBM X5 3950 单机
- 软件 SAP的HANA
- 结果
 - ▣ 记录查询 2毫秒
 - ▣ 三个大表连接操作, 9.2秒
 - ▣ 更新一天交易数据 59秒

| 表名称 | Records |
|------------|---------------|
| ETF表1 | 45,959 |
| ETF表2 | 570 |
| 产品基础信息表 | 97,456 |
| PUB_营业部关系表 | 368,132 |
| 持仓变化表 | 8,480,144 |
| 指定账户关系表 | 1,203,414,125 |
| 持仓数据表 | 4,412,295,708 |
| 申报数据表 | 969,151,941 |
| 成交数据表 | 963,307,302 |

| 计算机组件 | 参数 | 数量 | 备注 |
|-------|---|----|------------------------------|
| CPU | Intel(R) Xeon(R) CPU E7- 8870 @ 2.40GHz | 80 | 40 个核, Hyper Threads Enabled |
| 内存 | 1 TB | | 32 M 缓存 |
| 文件系统 | IBM GPFS 并行文件系统 | | 数据镜像 读写速度 ~633MB/s |

科学数据管理

46

- 无论是科学实验数据还是观测数据，需要领域专家，科学文献和设计数据专门用户
- 数据的标注和共享是长期的追求目标，涉及全链条的发展，data curation (数据监护)
- 所依赖硬件和系统架构与具体应用相关
- 研究问题：
 - ▣ 科学/统计数据库 (LNBL)
 - ▣ 阵列 (Array) 数据模型 (SciDB)

科学数据管理示例

47

□ 用于发现疾病抗体基因的工作流

- 跨多个数据集
(物理上分布)
- 工作流复杂
- 处理复杂

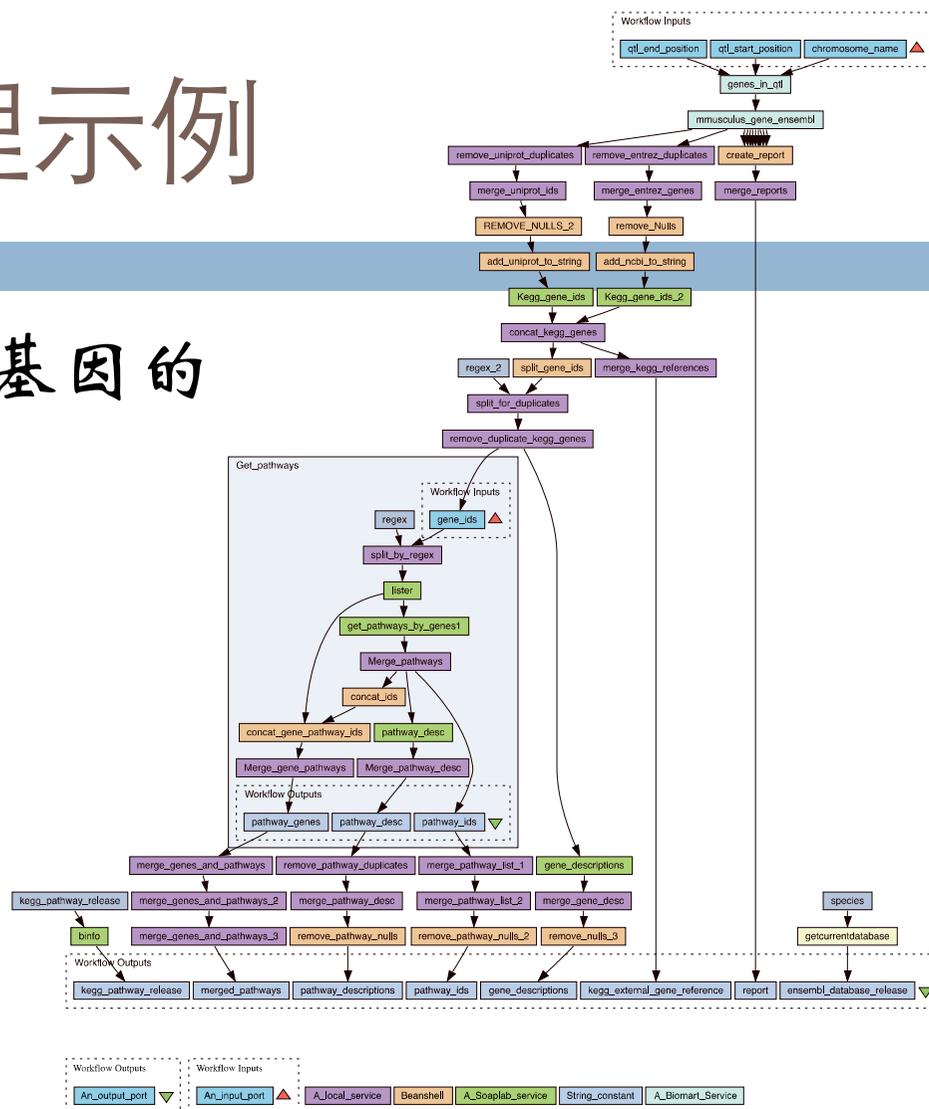
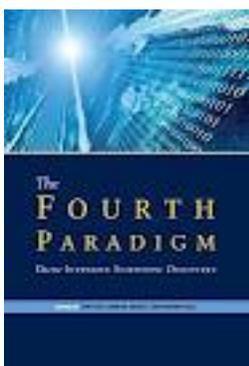


FIGURE 1.

A Taverna workflow that connects several internationally distributed datasets to identify candidate genes that could be implicated in resistance to African trypanosomiasis [11].

科学数据管理示例

48

- 可回溯的可视化分析
 - 即时可视化分析
 - 数据世系管理 (provenance)
 - 复杂的处理流程

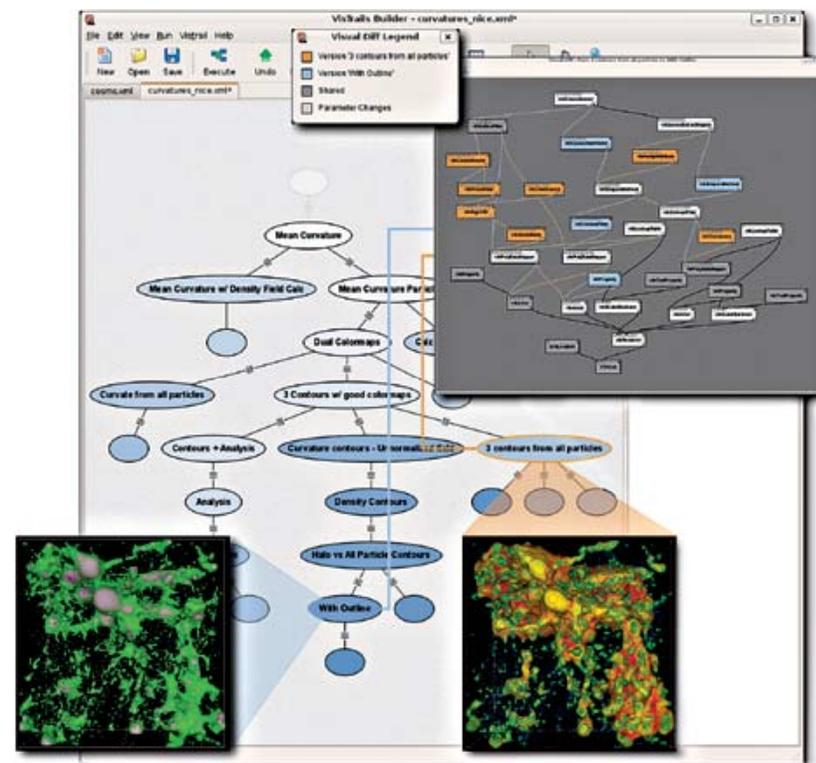
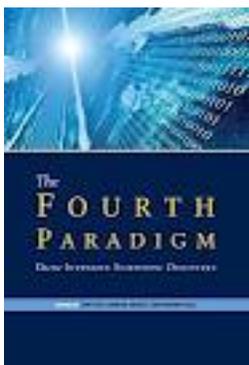


FIGURE 4. Representing provenance as a series of actions that modify a pipeline makes visualizing the differences between two workflows possible. The difference between two workflows is represented in a meaningful way, as an aggregation of the two. This is both informative and intuitive, reducing the time it takes to understand how two workflows are functionally different.

我们的机遇和实践

49

- Web数据（面向消费者）
 - 社交媒体数据分析
 - 计算广告
- 决策数据（面向企业）
 - 基于内存计算和新型系统架构的数据分析系统
 - 在重要的功能上大幅度提升性能
 - 研究内容：基准测试、可靠性、查询优化
- 科学数据（商业模式）
 - 基于位置的服务（LBS）（北斗、GPS、汽车导航）
 - 研究问题：移动对象数据管理、轨迹管理

值得探索的做法

50

- 应用驱动+企业合作
- 专注于具体领域应用
 - ▣ 舍弃 “One Size fits All”
 - ▣ 落实 “One Size Fits a Bunch”
- 及早开展Benchmark研究（执牛耳）
 - ▣ 数据分析、金融应用、Web数据
 - ▣ Workload (Performance, Scalability), Data Generator
- Big Data Appliance
- Open Source

提纲

- Hadoop的起源和历史
- Hadoop在数据管理中的地位
- 传统数据库发展回顾
- 什么是“大数据”
- 从数据库角度看大数据研究
- 结语

结语

- Hadoop和传统数据库一样，是大规模数据管理的一种有益的探索 **One Size Fits a Bunch**
- 大数据是一个笼统的概念，与应用密切相关，应该分类研究 **Web数据/决策数据/科学数据**
- 新的环境（数据采集、硬件技术、计算环境、应用需求）催生了大数据研究，但本质上而言：
大数据是数据库及相关学科的自然延伸
- 大数据研究是应用驱动的（制度、文化、习俗），
我们处在一个充满机遇的时间窗口
 - ▣ Enterprise-oriented Data Analytics Appliance
 - ▣ Consumer-oriented Location-Based Services Support
 - ▣ In-memory, Column Store, Big Data Appliance



敬请指正

谢谢！

[http://database.ecnu.edu.cn/
ayzhou@sei.ecnu.edu.cn](http://database.ecnu.edu.cn/ayzhou@sei.ecnu.edu.cn)