



武汉大学

WUHAN UNIVERSITY

TOTEM

珞珈图腾数据库实验室

PostgreSQL & DB Research@ 武汉大学

讲者：彭煜玮

合作者：彭智勇、宋伟、刘斌、
王黎维、李蓉蓉、杨先娣



实验室基本情况

- 教师队伍：彭智勇教授，副教授3名，讲师3名
- 博士生：12人，硕士生：12人
- 研究方向：
 - DBMS (PostgreSQL)
 - 对象代理数据库管理系统
 - 跨媒体数据按需管理
 - 专利数据管理与分析
 - Web社区管理与服务
 - 云计算环境下可信数据管理
 - 大数据背景下数据版权保护
- 教学：
 - 数据库内核分析
 - 数据库系统实现
 - 数据库原理
 - Oracle数据库
 - 数据挖掘
 - 操作系统
 -

PostgreSQL教学

- 组内多位老师承担《数据库原理》必修课程，面向计算机学院二年级本科生
- 时间：二年级第二学期，大约3月-6月，共 54 学时
- 课程规模：70-100人
- 教学方式：原理+实验（引导学生使用PostgreSQL）

PostgreSQL教学

- 彭智勇教授面向高年级本科生开设《数据库系统实现》选修课，
- 时间：每学年第一学期，大约9月-12月，共 54 学时
- 课程规模：50-100人
- 课程内容：DBMS实现原理
- 教学方式：全英文课程讲解+学生分析PostgreSQL印证课程内容

PostgreSQL教学

- 自2006年起在计算机学院硕士研究生一年级开设《数据库内核分析》课程
- 时间：每学年第一学期，大约9月-12月，共 54 学时
- 课程规模：15-25人
- 课程内容：PostgreSQL内核知识--进程结构、总体控制、存储管理、查询分析与优化、查询执行、并发控制等
- 教学方式：讲解+学生分析最新PostgreSQL特性

对象代理数据库管理系统 TOTEM（图腾）

研制背景

- 关系数据模型柔软性强，适合管理简单结构化数据，但难以为复杂非结构数据建模



- 对象数据模型语义丰富，适合复杂非结构化数据建模，但难以分割和重组对象



对象关系模型：

- ①数据难以统一建模
- ②各种机制难以协调
- ③臃肿庞大效率低下

代理对象的导入

关系数据库**柔软性强**
因为可通过关系代数自由地分割和重组

面向对象数据库**语义丰富**
因为数据和操作封装成对象，通过指针相互关联

代理对象



**柔软性强且语义丰富的
对象代理数据库**



对象可通过其代理
对象间接地进行分
割和重组

PostgreSQL与科研——自主数据库内核开发

- 2003年-2005年，在863计划项目支持下借鉴PostgreSQL的内部实现开发了新型数据库管理系统 TOTEM V1.0（图腾，对象代理数据库管理系统）
- 2006年至今，对TOTEM实现去PostgreSQL化，完成 TOTEM V2.0
- 2014年底至今，继续研发基于TOTEM 2.0的分布式数据库系统、研发TOTEM 2.0的复制机制、各类管理工具

对象代理数据库系统研制

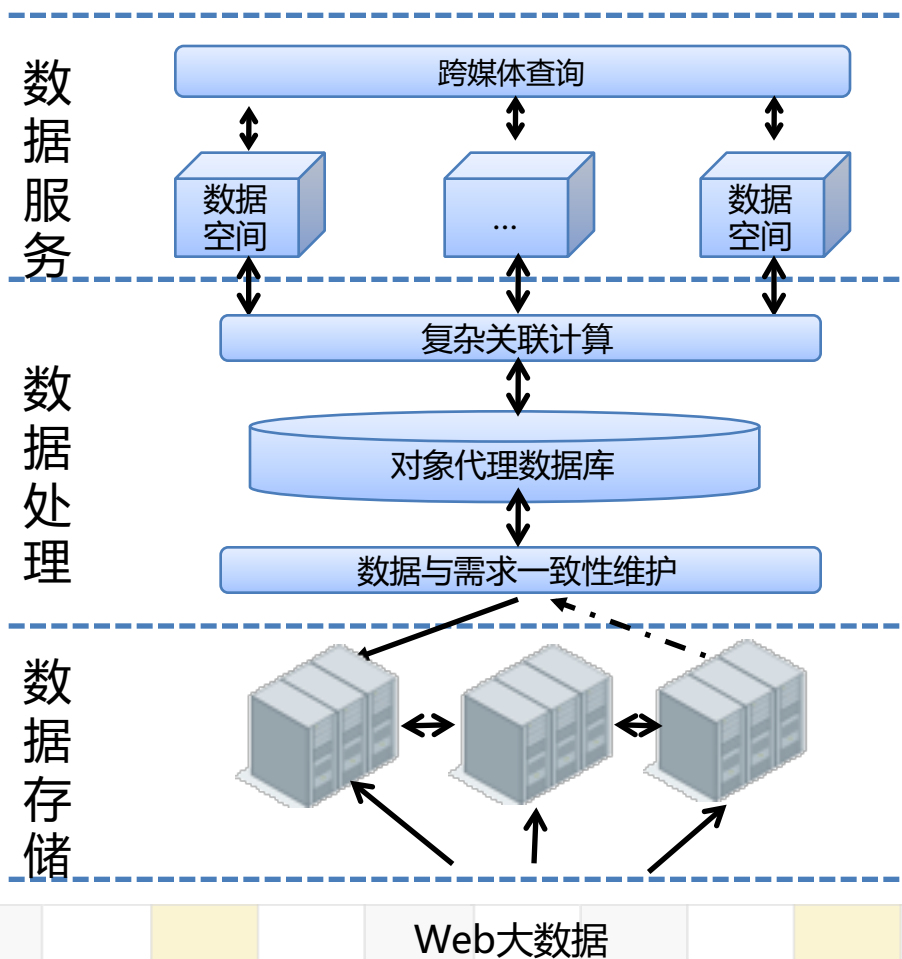
在国家863数据库重大专项课题“基于“关系+对象+代理”模型的对象管理技术研究”（2002AA4Z3450）资助下，研制了具有自主知识产权的对象代理数据库管理系统TOTEM，获教育部高等学校科技进步二等奖



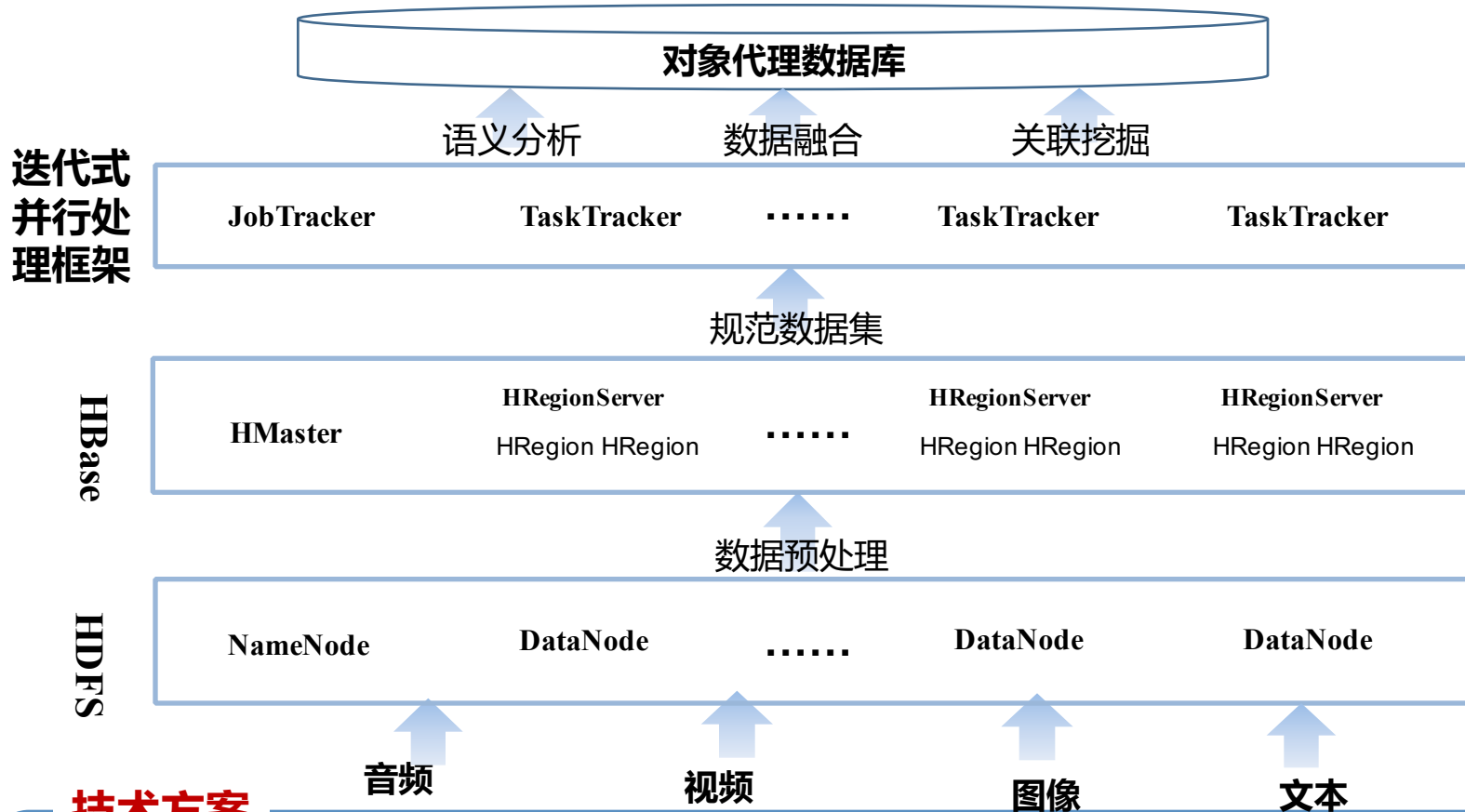
跨媒体数据按需管理

研究背景

- 建立用户个性化需求与Web大数据之间的动态联系，实现Web大数据的按需管理和基于复杂语义关联的跨媒体查询



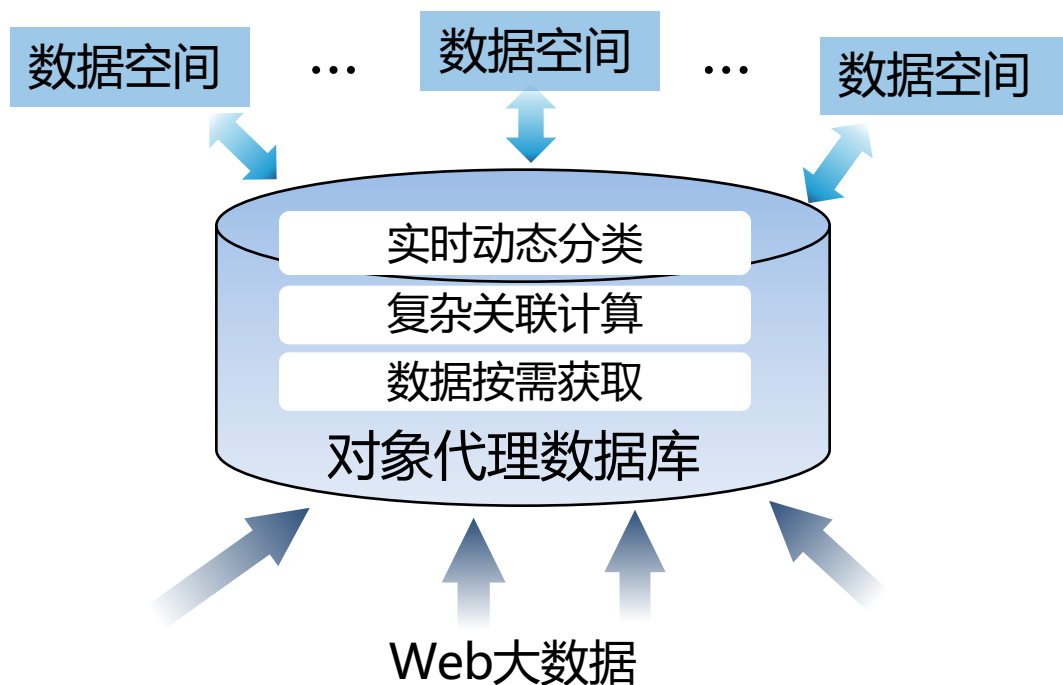
关键技术—海量数据存储



技术方案

1. 基于HDFS, HBase平台, 对音频、视频、图像、文本进行**分布式关联化**存储, 通过**重叠化、追加化**手段提高数据存储的局部性, 降低处理时网络开销
2. 采用**迭代式并行处理框架**实现海量复杂数据的语义分析、关联挖掘和数据融合

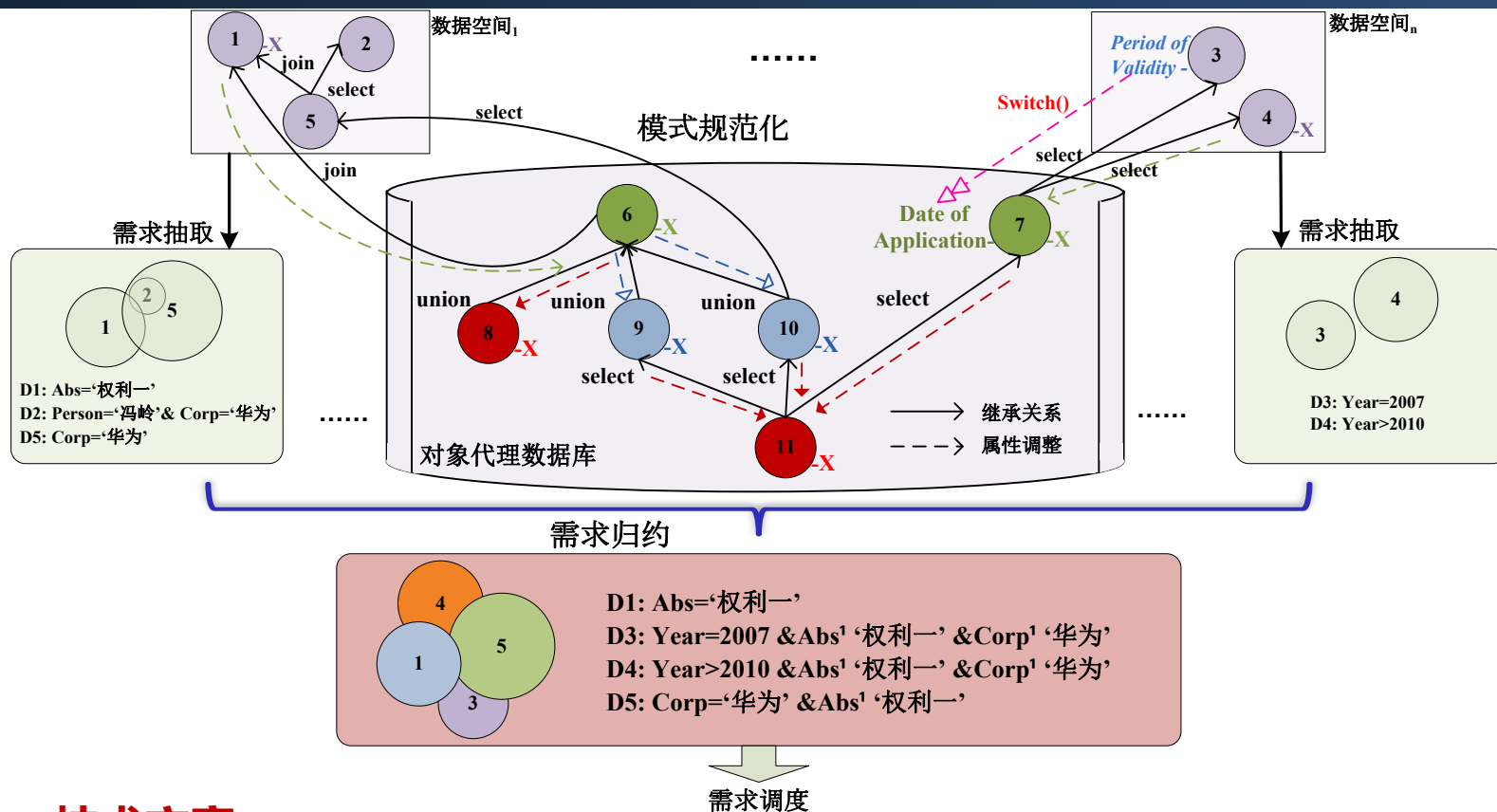
关键技术—复杂数据处理



技术方案

1. 基于对象代理数据库管理复杂语义关联，基于导航式跨类查询机制，实现海量复杂数据关联计算
2. 基于对象代理数据库更新迁移机制，依据实际应用需求，将数据动态分类到用户个性化数据空间
3. 根据应用需求重要度、数据库容量约束和基础数据变化规律等，设计数据按需获取优化调度算法

关键技术—数据空间管理



技术方案

1. 基于对象代理数据库创建数据空间，建立**对象代理数据库规范化理论**，设计新的范式和转换算法，消除模式中的属性不恰当定义、数据冗余和更新异常
2. 从数据空间中抽取个性化需求，运用**复合逻辑运算和推理**，分离需求重叠关系，实现公共需求归约化，消除冗余需求所带来数据按需获取与动态维护开销

TMUSIC: A Cross-Media Application

■ TMUSIC 主页面

专辑推荐

新歌推荐区

热门歌曲推荐

注册登录区

歌手新闻区

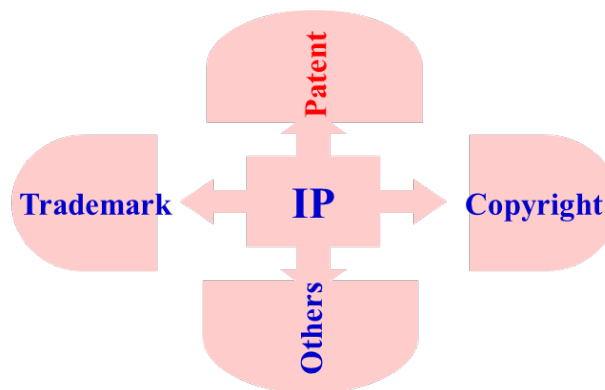
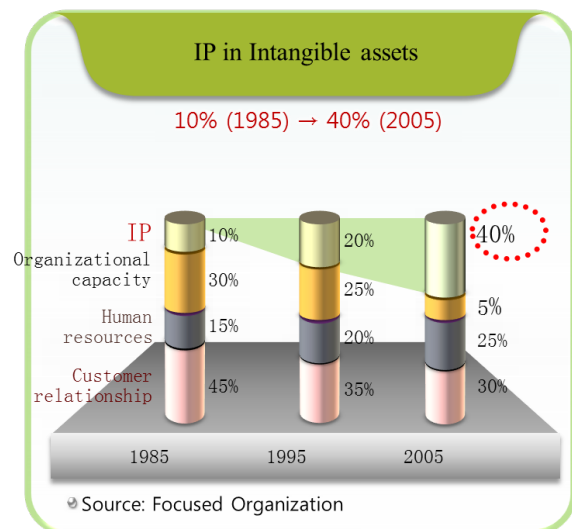
MTV推荐区

The screenshot shows the TMUSIC homepage layout. Annotations with arrows point to the following sections:

- 专辑推荐 (Album Recommendation):** Located at the top left, featuring a row of album covers.
- 新歌推荐区 (New Song Recommendation Area):** Located in the middle left, displaying a grid of song titles and artists with checkboxes and buttons for '全选' (Select All), '反选' (Inverse Select), and '加入列表' (Add to List).
- 热门歌曲推荐 (Popular Song Recommendation):** Located at the bottom left, showing a featured song '爱若走了' by 徐誉滕 with a list of other songs.
- 注册登录区 (Registration and Login Area):** Located at the top right, containing a '本站公告' (Site Notice) box and a '欢迎, demo' (Welcome, demo) box with links for '进入我的个性化库' (Enter my personalized library), '进入音乐盒' (Enter music box), and '退出' (Exit).
- 歌手新闻区 (Singer News Area):** Located in the middle right, titled '歌手新闻' (Singer News), listing recent news items about various artists.
- MTV推荐区 (MTV Recommendation Area):** Located at the bottom right, titled '热门MTV推荐' (Popular MTV Recommendation), showing video thumbnails and song information.

专利数据管理与分析

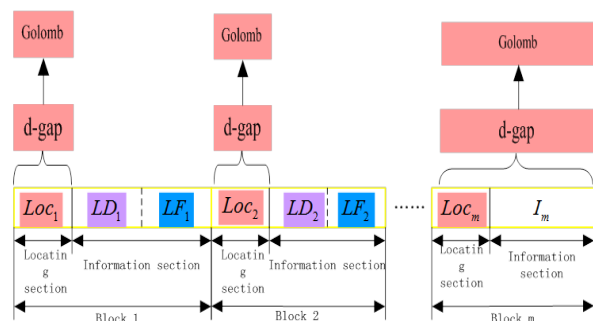
研制背景



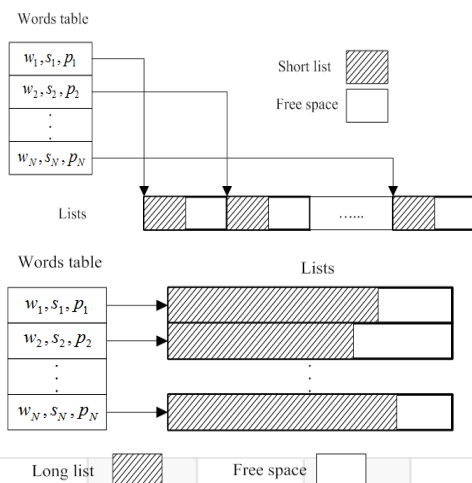
- 专利是一种非结构化的文本数据，包含了丰富的科技信息。
- 专利分析可以帮助企业
 - 避免重复和不必要的创新工作
 - 了解竞争对手的发展状况
 - 避免专利侵权

关键技术-专利管理

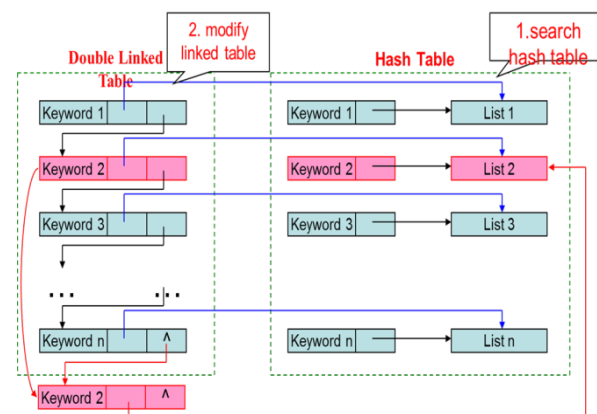
- 为了快速地检索专利信息，我们通常需要为专利文件建立索引。
- 在专利管理方面，我们的工作主要集中在
 - 索引压缩
 - 随机访问的分块倒排索引方法
 - 按块对倒排索引进行压缩
 - 索引构建和更新
 - 在构建和更新索引时，考虑长短列表区分开来，以进行预留空间分配
 - 查询优化
 - 提出最近未使用算法来管理内存中的查询结果缓存



索引压缩



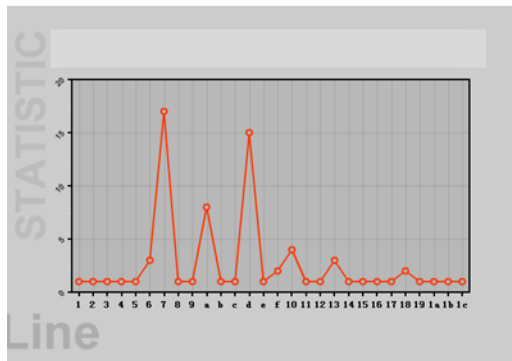
索引构建与更新



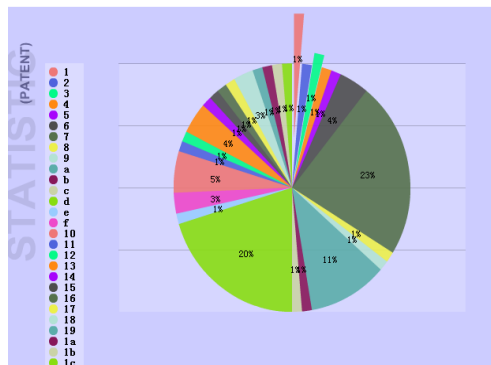
查询优化

关键技术-专利统计分析

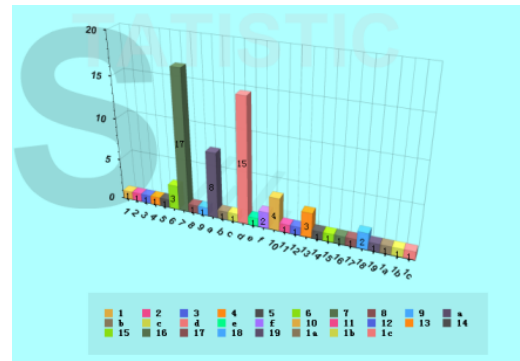
- 在统计分析方面，我们的工作主要集中在
 - 趋势分析
 - 分析关键技术的演化趋势
 - 在企业投资时提供决策支持
 - 技术分布分析
 - 识别竞争对手的技术分布情况
 - 企业技术对比
 - 比较两个企业的科技竞争力



趋势分析



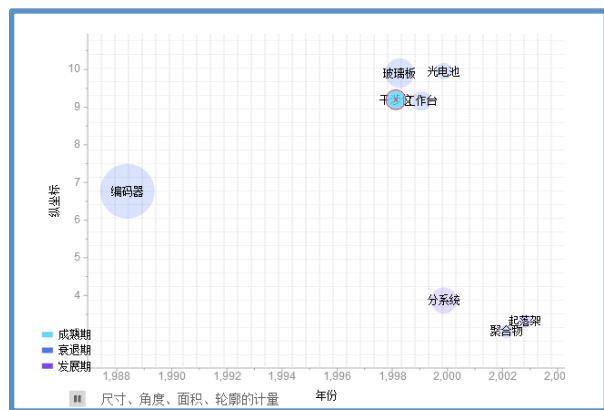
技术分布分析



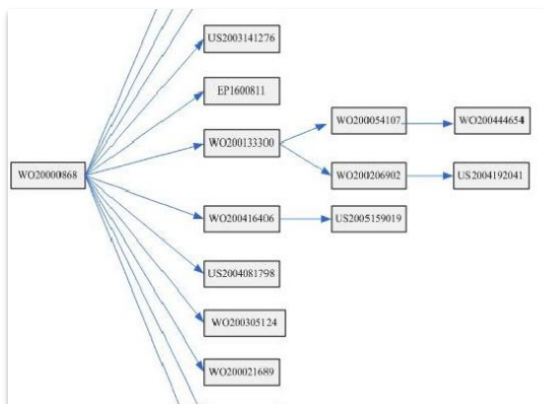
企业技术对比

关键技术-专利定性分析

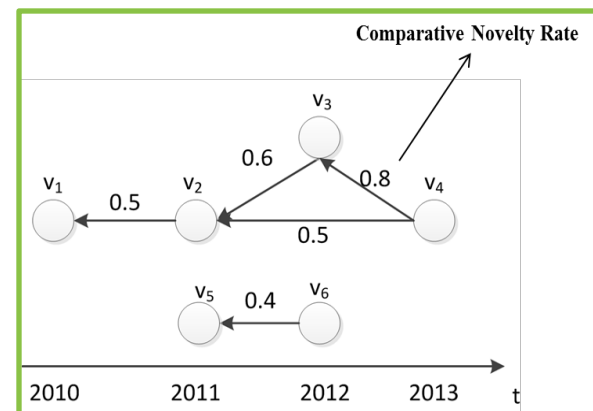
- 在定性分析方面，我们的工作主要集中在
 - 专利热点发现
 - 基于文本聚类的方法
 - 专利价值分析
 - 基于引用和专利发表时间的间隔
 - 新颖专利查找
 - 基于专利间的相对新颖度关联



专利热点发现



专利价值评估



新颖专利查找

Web社区管理与服务

研制背景



- 社交网络正在改变我们的生活Social Web is changing our life
 - 人们如何联系、交流和彼此合作等
- Web社区的理念
 - 由通过特定的社交媒介进行互动、为了共同的兴趣或目标、能跨地域和政治边界的个体构成的社交网络
- Web社区管理和服务
 - 基于对象代理数据库 (**Object Deputy Database**) 将Web社区作为复杂对象进行管理
 - 通过服务支持MOOC课程和团队合作

慕课社区管理

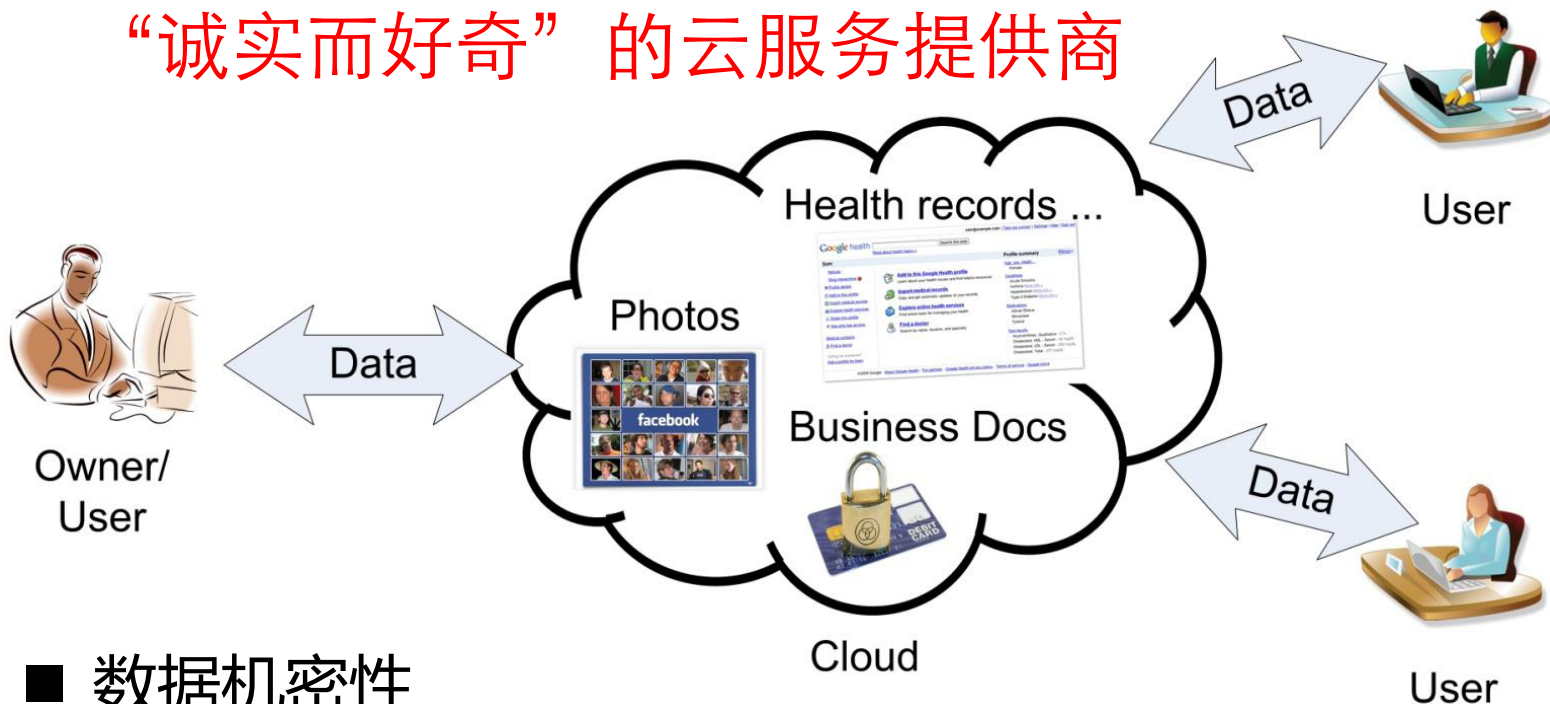
- 每一门MOOC课程被定义为一个慕课社区

PostgreSQL课程网站	社区主页	个人空间	课程组页	课件资料	在线论坛	管理员	退出
论坛新话题	我的资料库	社区论文					
留言: Mr....	上传资料	最新上传					
留言: Mr....	我上传的资料	最多下载					
留言: Mr....	创建资料库	qqq					
留言: Mr....	资料库列表	lll					
留言: Mr....	社区新鲜事	测试					
留言: Mr....	人在(2014-10-22 11:32:09)发帖! 请输入内容	cc					
留言: Mr....	人在(2014-10-22 11:32:05)发帖! 请输入内容	测试					
留言: Mr....	人在(2014-10-22 11:32:04)发帖! 请输入内容						
留言: Mr....	人在(2014-10-22 11:32:04)发帖! 请输入内容	声誉值排名					
留言: Mr....	人在(2014-10-22 11:32:04)发帖! 请输入内容	姓名	声誉值				
留言: Mr....	人在(2014-10-22 11:32:04)发帖! 请输入内容	1	王深港	3.5			
彭海平留言: 测试...	人在(2014-10-22 11:32:04)发帖! 请输入内容	2	李志浩	3.3			
张文杰留言: 请总控组的准备好PPT...	人在(2014-10-22 11:32:04)发帖! 请输入内容	3	张文杰	2.8			
熊胜超留言: 事务组关于开会,把时.....	人在(2014-10-22 11:32:04)发帖! 请输入内容	4	彭海平	1.3			
易晨辉留言: 分析该如何开始...	人在(2014-10-22 11:32:04)发帖! 请输入内容	5	罗永恩	1.2			
	人在(2014-10-22 11:28:59)发帖! 请输入内容	6	谭亚芳	0.9			
		7	谢森	0.8			
		8	蒋廉	0.8			
		9	刘正开	0.8			
		10	贺宏达	0.7			
		11	熊胜超	0.6			
		12	孙润康	0.6			

云环境下可信数据管理

云环境下数据安全需求

“诚实而好奇”的云服务提供商



- 数据机密性
 - 确保数据不被非法泄露、访问和使用
- 数据完整性
 - 确保云端数据是真实性和完整性
- 数据可用性
 - 确保云端数据的可用性及安全共享

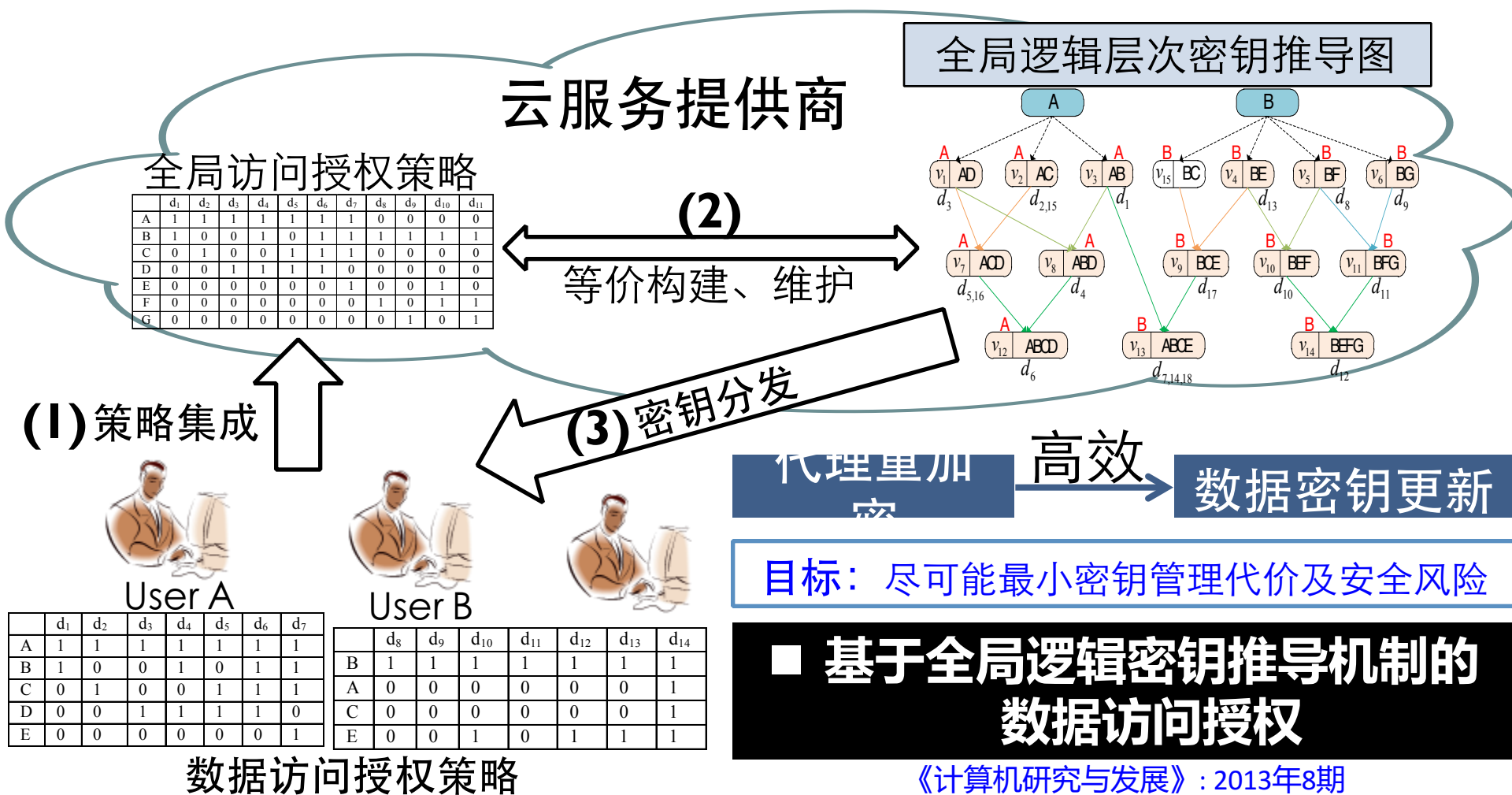
云环境下可信数据管理关键技术



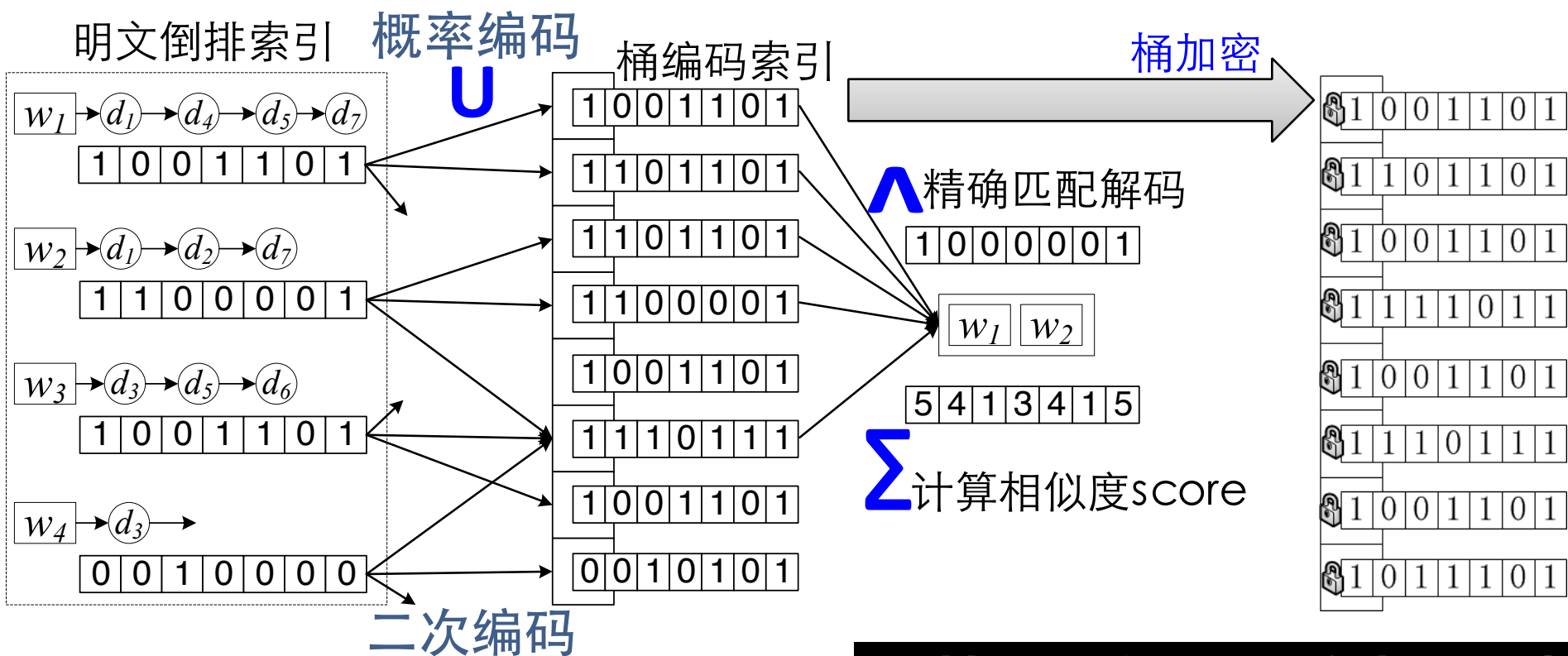
云端
数据
加密

- 如何进行安全高效的密文数据**访问授权**
- 如何进行安全高效的密文数据**查询操作**
- 如何进行安全高效的密文数据**正确性验证**

关键技术：密文数据访问授权



关键技术：密文数据查询操作

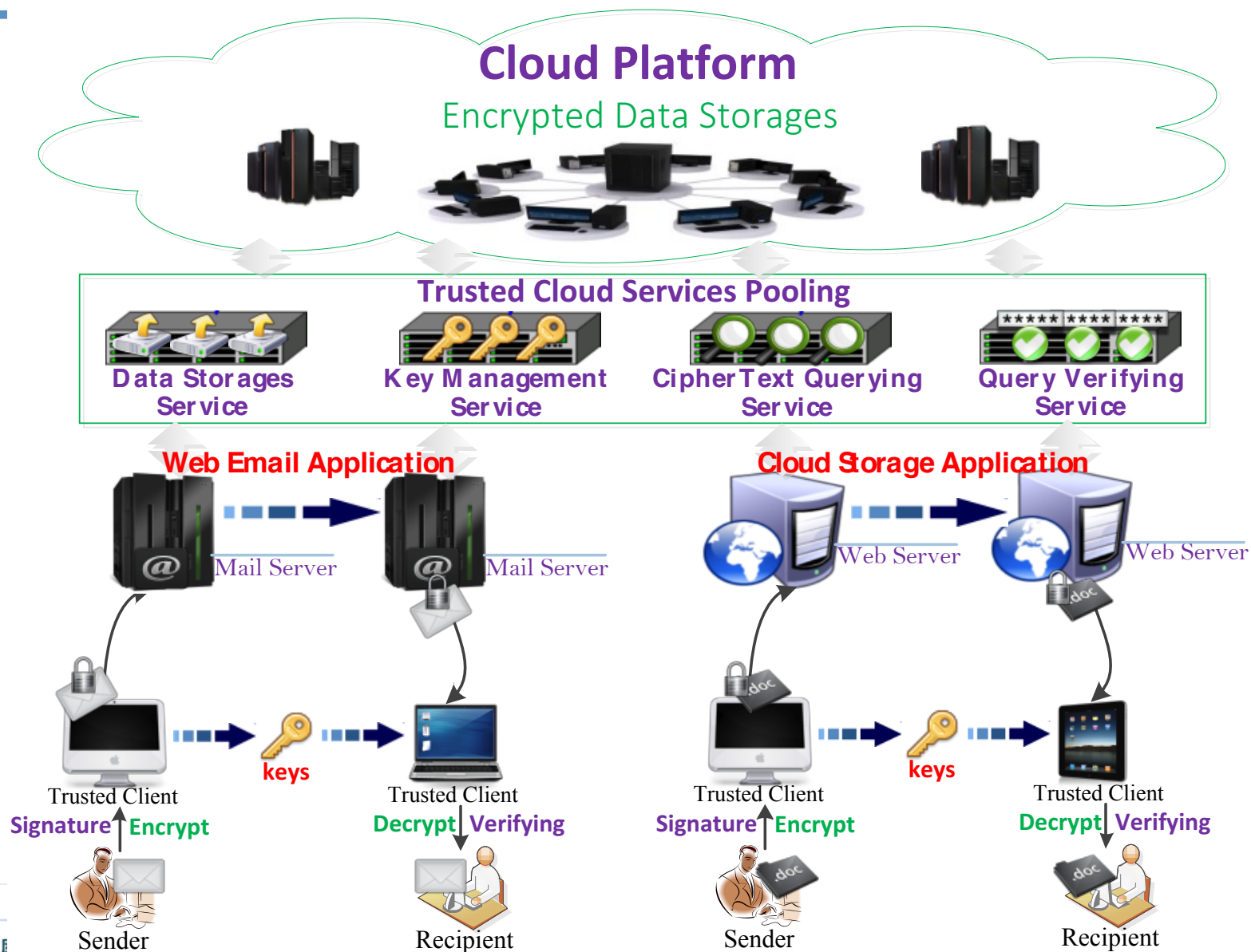


目标：基于单轮交互实现高效且隐私保护的多关键词查询操作

■ 基于桶编码的加密索引方法

《计算机学报》2012年10期；WAIM 2014 to appear

云环境下可信数据管理应用系统



云环境下可信数据管理应用系统

■ 可信云存储服务系统



安全密文文件共享

数据拥有者



被授权用户

云环境下可信数据管理应用系统

■ 可信Web电子邮件服务系统



基于关键词的
密文邮件检索

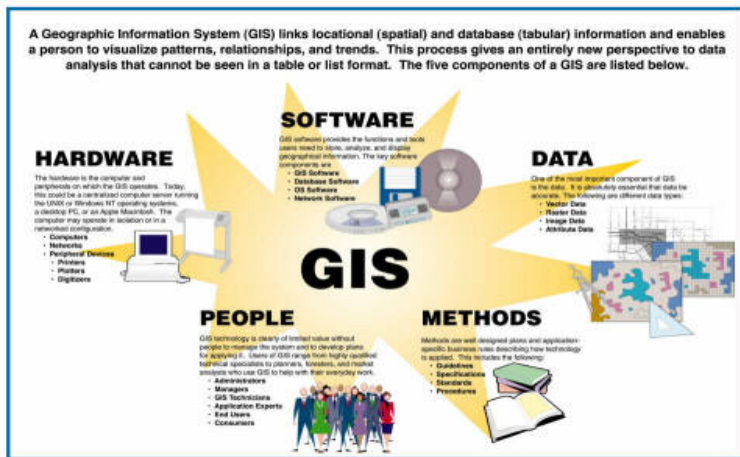
数据拥有者



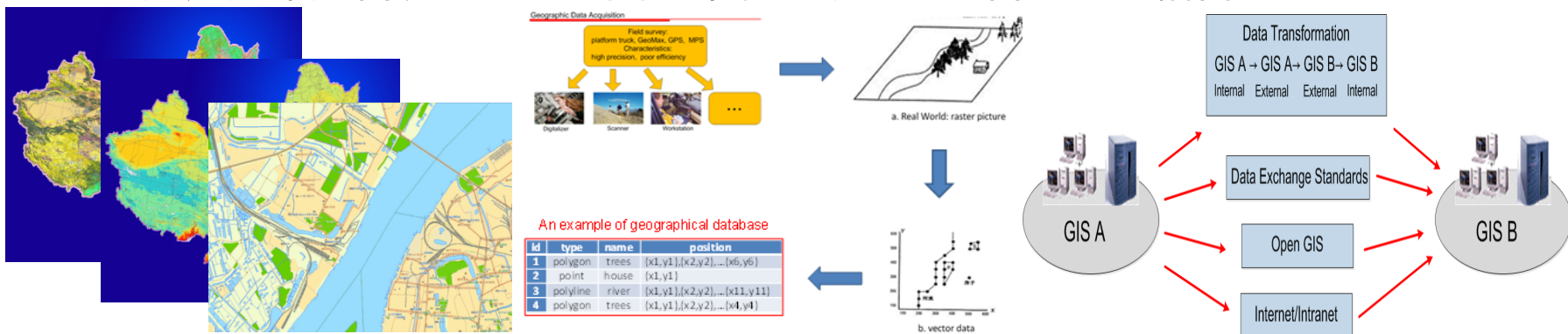
被授权用户

大数据背景下数据版权保护

研究背景与意义

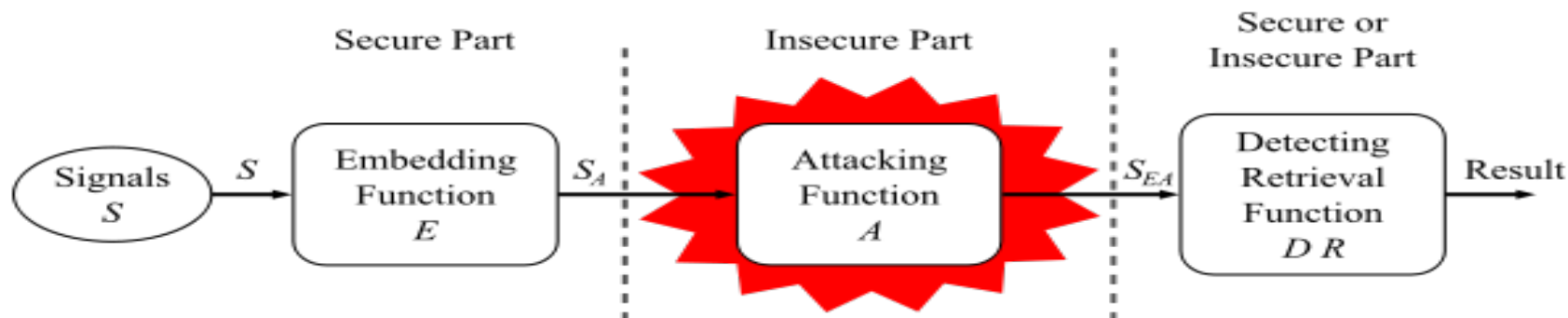


- 地理信息系统是将计算机硬件、软件、地理数据以及系统管理人员组织而成的对任一形式的地理信息进行高效获取、存储、更新、操作、分析及显示的集成
- GIS应用越来越普及：电子地图、车载导航、人口普查、地籍管理



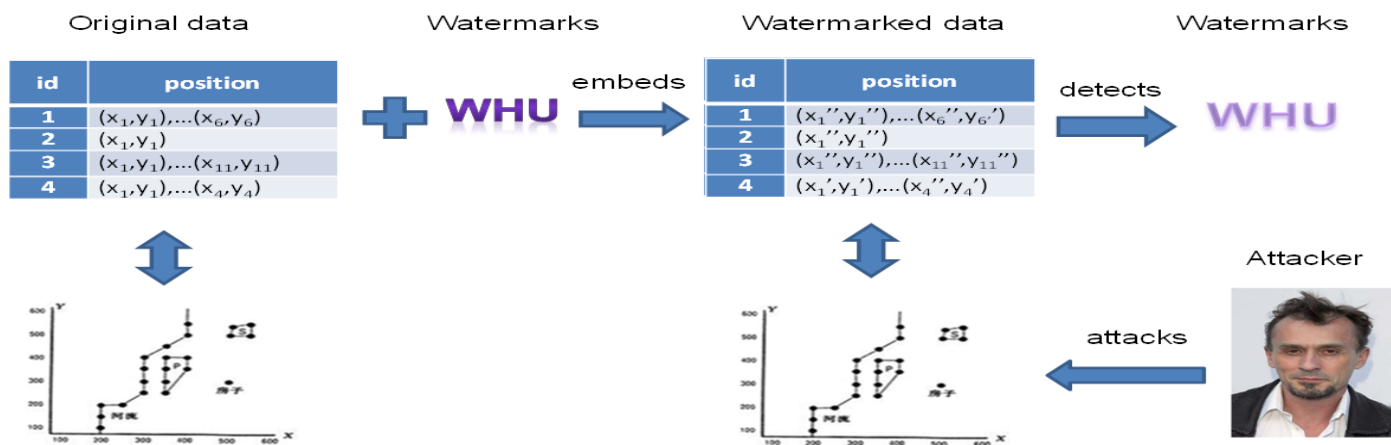
- ❑ 矢量地图是GIS的核心支持数据，地理数据的采集十分昂贵
- ❑ 地理数据的应用需要在不同的用户（组）间共享数据，版权、完整性需要得到保护

研究背景与意义



❑ 数字水印是最常用且最有效的数据版权保护方法之一

- 水印方法通过修改原始数据嵌入水印
- 检测时，水印可用于声明版权或验证数据完整性



❑ 本研究针对矢量地图进行版权保护与完整性验证

空间数据水印算法抗攻击性比较

算法	攻击方式					
	几何攻击	化简	插值	噪声	裁剪	地物插入
中国，武汉大学，2012	✓	✓	✓	✓	✓	✓
德国，卡昂大学，2010	✗	✓	✓	✓	✗	✗
希腊，亚里士多德大学，2007	✓	✗	✗	✓	✗	✗
IBM，阿尔马登研究中心，2003	✗	✗	✗	✓	✓	✓

目前国际上唯一可抵抗所有攻击的算法

自主知识产权

- 获得5项国家发明专利授权
 - 一种基于特征距离比值的地理数据库水印方法
 - 一种适合于polygon类型、高鲁棒的地理数据库水印方法
 - 一种抵抗几何攻击的矢量地图水印方法
 - 基于网格的矢量空间数据数字水印方法
 - 基于数字水印的保护地理数据版权的交易方法
- 获得1项软件著作权
 - 珞印地理数据水印工具



国际先进水平

在国家863计划地球观测与导航技术领域2006年度专题《空间数据水印技术及其服务体系的研究与开发》项目资助下，研制了具有自主知识产权的珞珈地理数据版权保护水印系统。经湖北省科技厅组织的成果鉴定会鉴定认为总体达到国际先进水平。



系统展示

主界面上的大部分功能都能在相应的菜单项中找到



地图装载/卸载

水印嵌入/检测控制

备份方式

水印类型

水印算法选择

主要负责选中要处理的地图对象，显示已经加载到工作区的地图

地图信息和高级属性：显示地图中地物数、半径、顶点数等信息；

地图显示区及辅助工具栏：用于地图显示、漫游、缩放视图、查询、鹰眼等功能

地图平移

地图旋转

地图缩放

地图化简

地图噪声

地图裁剪

地物修改

地物增加

地物删除

应用推广

- 武汉吉嘉伟业科技发展有限公司
- 湖北省测绘局
- 国家安全局
- 测绘遥感信息工程国家重点实验室

用户意见

随着GIS技术的广泛应用，GIS应用系统中的空间信息数据安全问题日益重要。如何保护空间信息的版权？数字水印技术为数据版权保护问题提供了最有力和最有效方案。数字水印可以看成是加在原始数据中的不可见或不可感知的数据信号，并且，可以借助相关保护对原始数据中的信号进行提取。自从数字水印技术提出以来，出现了很多不同类型的旨在保护空间数据版权的空间数据水印算法，比如构造影像水印算法和矢量数据水印方法。

测绘遥感信息工程国家重点实验室王作刚教授在上海嘉利软件项目以及湖北省交通厅地理信息支撑系统等项目中所用国家863项目“空间数据水印技术及其服务体系的研究与开发”成果，对空间信息海量数据进行水印加密，该技术能够有效抵抗攻击，保护数据，加密水印的加密数据效果良好，特此证明。

武汉大学测绘遥感信息工程国家重点实验室
2011年10月3日



矢量数据水印技术应用意见

随着科学技术的进步，GIS的应用已经进入我们的生活和生产中，成为GIS应用的主要内容之一。空间数据的价值不言而喻，随着计算机技术和网络技术的飞速发展，空间数据通过网络的传输和共享比以往任何时候都方便得多，对空间数据的非法复制和分发在网络上越来越普遍，这给网络数据的版权保护带来了极大的挑战。要建立一个安全的环境，需要许多关键技术的支持，比如数字水印技术、密钥管理等。

我公司利用国家863项目“空间数据水印技术及其服务体系的研究与开发”成果，对1:100, 1:2000和1:50000的矢量数据产品进行水印加密，加密数据水印加密有效，抗攻击性强，使用方便，加密后数据仍然保持原有精度和符合国家规定的范围内，符合空间数据使用要求，该技术在项目应用中效果良好，特此证明。

武汉吉嘉伟业科技发展有限公司
2011年6月16日



谢谢! & 提问!