大数据经济学 文献总结

MIT-Plato 团队

1. 俞立平.大数据与大数据经济学**[J].**中国软科学，**2013**，**7:018.**

摘要：本文从大数据的发展现状分析入手，讨论了大数据对传统经济学的挑战，首次提出大数据经济学的概念。认为大数据经济学包括大数据计量经济学、大数据统计学和大数据领域经济学，并分析了大数据经济学与信息经济学、信息技术等相关学科的关系，最后对大数据经济学发展前景进行了展望，认为大数据经济学不仅将理论科学、实验科学、复杂现象模拟统一在一起，而且将自然科学和社会科学统一在一起，将理论研究与实践应用实时地统一在一起，大数据经济学具有"智能经济学"的特点。

2. 王新才，丁家友.大数据知识图谱:概念，特征，应用与影响**[J].**情报科学，**2013**，**9:01.**

摘要：近年来"大数据"的理念与实践活动迅速拓展和渗透到社会的各个领域。本文采用可视化工具 Gephi，借助知识图谱的理念，在文献调研的基础上，初步探索大数据的发展脉络、应用范围以及对人类社会产生的影响。

3. 钱文静， 邓仲华.云计算与信息资源共享管理**[J].**图书与情报，**2009**，**4(3).**

摘要：云计算自从被提出以来，飞速发展，目前已经有了一些可用的云计算服务。云计算受广泛的推崇，是因为它可利用最小化的客户端实现复杂高效的处理和存储的特点，这给我们带来巨大的发挥空间。若将云计算网络推广到信息资源共享管理中，无疑会极大的提高信息资源共享的能力。文章在介绍云计算概念、特点及最新发展的基础上，以我国高校信息资源共享为例，探讨了云计算技术与方法为改善高校信息资源共享能力带来的变革，并试图探寻建设起校际间云计算网络提高信息资源利用率的方式。最后，讨论了校际数据库共享云计算方案。

4. 赵华茗，李春旺，李宇.云计算及其应用的开源实现研究**[J].**现代图书情报技术，**2009**，**9:1-6.**

摘要：对比分析当前有代表性的云计算参与企业所采用的云计算实现方案和技术特点，总结云计算是一种能够提供动态资源池、虚拟化和高可用性的计算平台，包含两个方面的含义:底层的基础设施平台和构建在这个平台之上的应用程序。通过集成整合 Eucalyptus 开源云计算平台系统和 Liferay 开源内容管理系统，剖析具有特定应用的云计算平台构建过程，并着重介绍在基于虚拟技术的云计算平台中，使用 Xen 虚拟技术制作基于 Linux 系统的可定制服务的实例映像文件，使用 AMI 工具运行制作的实例(虚拟服务器)和部署应用服务等技术要点，使读者能够明确掌握云计算平台的搭建过程及云计算实例应用的快速部署与迁移。

5. 李晨晖，张兴旺，崔建明，陈超泉.一种基于异构云计算平台的资源管理模型**[J].**情报理论与实践.**2013(01)**

摘要：文章将多 Agent 与云计算技术相融合，提出一种基于异构云计算平台的资源管理模型，即利用多 Agent 技术特性，在异构云计算环境中构造相应的资源、Agent，通过多 Agent 的交互协作，实现云计算环境中的资源调度的有效分配及管理，实现现有云计算平台之间的数据和应用的互操作性和可移植性，并以 Chukwa 为例，验证该模型在现有技术条件下的可行性及其在性能上的优势。

6. 洪程.国外科学数据服务现状研究**[J].**图书馆杂志.**2012(10).**

摘要：科学数据服务是充分利用科学数据和满足科研需求的必要环节与手段，国外图书情报机构的科学数据服务已取得一定的成绩。从数据存储与发布服务、数据发现与获取服务、数据管理规划服务、数据分析服务、数据引用服务和用户社区服务来分析其服务现状，总结在服务理念、服务内容与方式、服务支撑上的特点。

7. 蒋颖.欧洲社会科学数据的服务与共享**[J].**国外社会科学**.2008(05)**

摘要：社会科学数据对于社会科学领域的实证研究具有重要价值。本文以英国、德国和荷兰等国为例介绍了欧洲的社会科学数据管理和服务情况。欧洲各国很早就意识到进行统一的数据管理、服务和共享的重要性，并开始进行这方面的实践。最主要的措施有两项:一是建立国家级的数据档案机构，集中管理和保存各种社会科学数据，在一定范围内进行共享;另一个是利用现代化的手段，在保证不泄密的基础上，逐步开放国家统计局的微观数据，以供研究使用。此外，这些国家还开发了相关的数据服务软件 Nesstar，以方便数据管理和用户检索。

8. 陈康，郑纬民**.**云计算**:**系统实例与研究现状**[J].**软件学报**.2009(05).**

摘要：针对云计算这样一个范畴综述了当前云计算所采用的技术，剖析其背后的技术含义以及当前云计算参与企业所采用的云计算实现方案.云计算包含两个方面的含义:一方面是底层构建的云计算平台基础设施，是用来构造上层应用程序的基础;另外一方面是构建在这个基础平台之上的云计算应用程序.主要是针对云计算的基础架构的研究与实现状况给出综述，对于云计算的应用也有所涉及.云计算有3个最基本的特征:第1个是基础设施架构在大规模的廉价服务器集群之上;第二是应用程序与底层服务协作开发，最大限度地利用资源;第3个是通过多个廉价服务器之间的冗余，通过软件获得高可用性.云计算达到了两个分布式计算的重要目标:可扩展性和高可用性.可扩展性表达了云计算能够无缝地扩展到大规模的集群之上，甚至包含数千个节点同时处理.高可用性代表了云计算能够容忍节点的错误，甚至有很大一部分节点发生失效也不会影响程序的正确运行.通过此文可以了解云计算的当前发展状况以及未来的研究趋势.

9. 罗军舟，金嘉晖，宋爱波，东方**.**云计算**:**体系架构与关键技术**[J].**通信学报**.2011(07)**

摘要：系统地分析和总结云计算的研究现状，划分云计算体系架构为核心服务、服务管理、用户访问接口等3个层次。围绕低成本、高可靠、高可用、规模可伸缩等研究目标，深入全面地介绍了云计算的关键技术及最新研究进展。在云计算基础设施方面，介绍了云计算数据中心设计与管理及资源虚拟化技术;在大规模数据处理方面，分析了海量数据处理平台及其资源管理与调度技术;在云计算服务保障方面，讨论了服务质量保证和安全与隐私保护技术。针对新型的云计算应用和云计算存在的局限性，又探讨并展望了今后的研究方向。最后，介绍了东南大学云计算平台以及云计算研究与应用方面的相关成果。

10. 黎春兰，邓仲华**.**论云计算的价值**[J].**图书与情报**.2009(04)**

摘要：云计算一开始就立足于商业，为企业提出了一个更为广阔的发展空间。文章通过介绍各主流厂商比如 Google，IBM，微软等的云计算的理念及云计算的共有特点，从内、外部架构来分析云计算的价值，最后提出了云计算模式所面临的关于安全性、竞争性等的挑战。

11. 郑湃，崔立真，王海洋，徐猛**.**云计算环境下面向数据密集型应用的数据布局策略与方法**[J].**计算机学报**.2010(08)**

摘要：云计算环境下面向流程的数据密集型应用已被广泛应用于多个领域.面对多数据中心的云计算环境，这类应用在数据布局方面遇到了新的挑战，主要表现在如何减少跨数据中心的数据传输、如何保持数据间的依赖性以及如何在提高效率的同时兼顾全局的负载均衡等.针对这些挑战，文中提出一种三阶段数据布局策略，分别针对跨数据中心数据传输、数据依赖关系和全局负载均衡三个目标对数据布局方案进行求解和优化.实验显示，文中提出的数据布局策略具有良好的综合性能，特别是在降低流程执行过程中由跨数据中心数据传输所导致的时间开销方面，效果尤为明显.

12. 刘正伟，文中领，张海涛**.**云计算和云数据管理技术**[J].**计算机研究与发展**.2012(S1)**

摘要：随着各种新技术的发展，企业的关键信息以几何级速度增长，更多的数据需要保存更长的时间.伴随着云计算技术的发展，云计算已经成为一种全新的互联网应用模式.而在云计算对海量的数据高效管理，云端数据精确精准快速查询成为越来越重要的问题.一个新的面向云计算的数据管理研究领域正逐渐形成，在云计算技术的基础上，提出了云数据管理的概念.分析 GFS，BigTable，Dynamo 等当前互联网主流云数据管理系统的基本原理，并针对未来云数据管理架构进行分析，最后指出了云数据管理领域的主要研究方向.

13. 张兴旺，李晨晖，秦晓珠**.**云计算环境下大规模数据处理的研究与初步实现**[J].**现代图书情报技术**.2011(04)**

摘要：将云计算技术引入到大规模数据处理过程中，提出在集中或分布管理的大量廉价计算机集群上构建动态的、可扩展的、高性价比的、易使用的高性能计算平台，创建一种基于云计算的大规模数据处理的框架模型。论述在这一环境下的大规模数据处理的方法和应用，通过搭建相应的计算平台，验证计算机集群及框架模型的可行性。

14. 张兴旺，李晨晖，秦晓珠**.**基于云计算的数字化信息资源建设模型的研究**[J].**情报理论与实践**.2011(08)**

摘要：针对云计算技术在数字化信息资源建设体系中的应用需求，提出适用于高校图书馆的中小型云计算中心的规划建设模型;通过对云计算技术的研究，结合数字化信息资源建设的业务特征，构建了基于云计算的数字化信息资源建设体系架构模型、实施方案，同时通过对 MapReduce 架构的核心思想和开发技术的研究，初步实现了"数据云"服务，完善和验证了这一体系架构模型。

15. 罗先觉，尹锋林**.**云计算对知识产权保护的若干影响**[J].**知识产权**.2012(04)**

摘要：云计算作为变革性的信息技术，已对知识产权保护和若干知识产权制度产生了重要影响。与本地计算模式相比，云计算一方面可以更有效地防止著作权侵权，另一方面需要依赖于合同、技术措施、商业方法专利等手段提供综合性的补充保护。云计算会对合理使用原则、发行权穷竭原则构成挑战。云计算环境下，用户的商业秘密保护可能因为数据迁移困难而受到妨碍。由于云计算尚处于发展初期，还不能认为知识产权制度需要进行重大修改。

16. 何大义**.**熵在数据分析中的应用研究**[J].**统计与决策**.2005(08)**

摘要：本文从极大熵准则出发，在信息熵定义的基础上给出了数据的等值性、等值度、等偏性以及等偏度的定义，并讨论了这些概念的一些基本性质及其应用，为数据分析提供了一种新的思路和方法。

17. 杨再高，智慧城市发展策略研究**[J].**科技管理研究**.2012(07)**

摘要：2008年 IBM 提出智慧城市理念及愿景后，建设智慧城市很快成为世界各国及城市在金融危机后寻求新一轮发展的重大战略举措。智慧城市不仅能引发新一轮技术创新及物联网产业发展，促进城市经济社会发展智慧化，而且能促进城市建设管理智慧化，使人类以更加精细和动态的方式管理城市生产与生活，为居民创造更加美好的城市生活。简要介绍智慧城市产生背景、内涵和特征，分析对城市发展产生的影响和面临问题，提出包括加快智慧城市基础设施建设、物联网技术研发及产业、社会发展和城市管理智慧化、重点领域应用示范等发展策略，为智慧城市赢在未来夯实基础。

18. 逄金玉**.**“智慧城市”——中国特大城市发展的必然选择**[J].**经济与管理研

究.**2011(12)**

摘要：由信息化、数字化到智慧城市的发展，相应要求特大城市积极向智慧城市转变，这也是特大城市解决突出问题，适应管理复杂化的必然要求。本文在分析特大城市与智慧城市关系的基础上，分析了智慧城市的主要特征，并以上海市与宁波市的智慧城市建设为例，展示了智慧城市的技术共性与相异的应用路径。在技术、体制、应用、产业促进等方面提出了推动特大城市向智慧城市发展的建议。

19. 史璐.智慧城市的原理及其在我国城市发展中的功能和意义**[J].**中国科技论坛.**2011(05)**

摘要：本文从智慧地球的理念入手，系统论述智慧城市的理论基础、概念和特征;分析智慧城市的业务构架和具体功能;前瞻性地论述智慧城市建设在我国城市发展中的重要意义;最后，本文对智慧城市建设应注意的问题以及配套改革措施进行了展望。论文在系统综述的基础上，力求深入浅出、论述严谨，供政府相关决策者和研究人员参考。

20. 陈博，高光耀.智慧城市的建设路径、核心和推进策略研究**[J].**管理现代化.**2013(01)**

摘要：智慧城市被认为是继数字城市和智能城市建设之后城市信息化的高级阶段，并已成为国内众多城市未来发展的重要战略选择。本文从分析智慧城市内涵入手，提出我国智慧城市分"三步走"的建设路径构想，并在系统科学和复杂性科学视角下提出智慧城市建设的核心和推进策略。

21. 张千帆，梅娟.竞争平台的移动数据业务定价研究**[J].**图书情报工作.**2010(16)**

摘要：围绕移动数据业务平台的双边市场定价这一核心问题，构建竞争平台下的定价策略系统框架，建立竞争平台定价的基本模型，并探讨竞争平台的移动数据业务定价方式，包括注册费、交易费、两步制收费，并分析定价与成本、网络外部性因素、平台差异化程度、匹配技术、交易量等的关系，最后对网络运营商提出服务定价建议。

22. 李晨晖，崔建明，陈超泉**.**大数据知识服务平台构建关键技术研究**[J].**情报资料工作**.2013(02)**

摘要：文章分析了大数据知识服务模式的运行机理，建立了大数据知识服务平台构建体系架构，阐述了大数据知识服务平台构建过程中所涉及的主要关键技术，并对实施大数据知识服务模式的思路和发展提出了建议。

23. 文小平**.**通用数据处理平台的研究与设计**[J].**中国统计**2012(11)**

摘要：自接触以数据库为后台的数据处理软件以来，见证、经历并使用了统计上所有的专业程序、普查程序，通过对金碟、新中大、清华同方等 ERP 软件和各类库存软件、数据采集软件的分析和使用，明白了统计系统使用的数据处理软件最具代表性，也最复杂。

24. 刘金玲**.**基于商业平台的信息资源整合研究**[J].**图书馆理论与实践**.2012(07)**

摘要：以元数据描述框架为支撑的信息集成、检索、资源共享等信息服务系统正在逐步形成。图书馆界应该转变观念，突破将资源共享局限在图书馆内部事务的观念，以社会化服务为依托，利用商业整合平台，使用元数据组织方式将信息资源的元数据进行整合，构建科学、合理、实用的用户信息体系架构，实现对信息资源目录的创建、检索、更新以及权限管理，实现信息资源的有效整合。

25. 毛定祥**.**经济数据挖掘**[J].**上海大学学报**(**社会科学版**).2006(02)**

摘要：数据挖掘是从数据库中发现经济知识模式的有效工具，应当把数据挖掘的方法技术引入和应用到实证经济分析中去，以促进经济学的现代化、科学化，提高经济决策水平。

26. 王世伟**.**说**“**智慧城市**”[J].**图书情报工作**.2012(02)**

摘要：分析智慧及智慧城市的性质、特点和发展阶段，着力从创新未来城市发展的顶层设计、因城而宜地推进城市的可持续发展、加强和创新社会管理的重要路径、构筑面向市民的泛在服务、提供新兴产业发展的动力引擎等多角度论述智慧城市带给人们的启示。

27. 刘云，刘东苏，周津慧**.**数据仓库中的数据组织与数据有效性分析**[J].**情报理论与实践**.2000(04)**

Abstract: The organization structure of the data in a data warehouse is discussed. The effectiveness of the data is analyzed. In view of the stale data and ineffective data in a data warehouse, this article describes the causes for their occurrence and puts forward a solution.

28. Bughin J, Chui M, Manyika J. Clouds, big data, and smart assets: Ten tech-enabled business trends to watch[J]. McKinsey Quarterly, 2010, 56(1): 75-86.

Abstract : Advancing technologies and their swift adoption are upending traditional business models. Senior executives need to think strategically about how to prepare their organizations for the challenging new environment.

29. Herodotou H, Lim H, Luo G, et al. Starfish: A Self-tuning System for Big

Data Analytics[C]. CIDR. 2011, 11: 261-272.

Abstract: Timely and cost-effective analytics over ñBig Dataò is now a key ingredient for success in many businesses, scientific and engineer-ing disciplines, and government endeavors. The Hadoop software stackð which consists of an extensible MapReduce execution engine, pluggable distributed storage engines, and a range of procedural to declarative interfacesð is a popular choice for big data ana-lytics. Most practitioners of big data analyticsð like computational scientists, systems researchers, and business analystsð lack the expertise to tune the system to get good performance. Unfortunately, Hadoopôs performance out of the box leaves much to be desired, leading to suboptimal use of resources, time, and money (in pay-as-you-go clouds). We introduce Starfish, a self-tuning system for big data analytics. Starfish builds on Hadoop while adapting to user needs and system workloads to provide good performance automatically, without any need for users to understand and manipulate the many tuning knobs in Hadoop. While Starfishôs system architecture is guided by work on self-tuning database systems, we discuss how new analysis practices over big data pose new challenges; leading us to different design choices in Starfish.

30. Weiss A. Computing in the clouds[J]. networker, 2007, 11(4).

Abstract: To some, the cloud looks like Web-based applications, a revival of the thin-client. To others, the cloud looks like utility computing, a grid that charges metered rates for processing time. Then again, the cloud could be distributed or parallel computing, designed to scale complex processes for improved efficiency. Maybe everyone is right. There are many shapes in the clouds

31. Dikaiakos M D, Katsaros D, Mehra P, et al. Cloud computing: Distributed Internet computing for IT and scientific research[J]. Internet Computing, IEEE, 2009, 13(5): 10-13.

Abstract: Cloud computing is a disruptive technology with profound implications not only for Internet services but also for the IT sector as a whole. Its emergence promises to streamline the on-demand provisioning of software, hardware, and data as a service, achieving economies of scale in IT solutions' deployment and operation. This issue's articles tackle topics including architecture and management of cloud computing infrastructures, SaaS and IaaS applications, discovery of services and data in cloud computing infrastructures, and cross-platform interoperability. Still, several outstanding issues exist, particularly related to SLAs, security and privacy, and power efficiency. Other open issues include ownership, data transfer bottlenecks, performance unpredictability, reliability, and software licensing issues. Finally, hosted applications' business models must show a clear pathway to monetizing cloud computing. Several companies have already built Internet consumer services such as search, social networking, Web email, and online commerce that use cloud computing infrastructure. Above all, cloud computing's still unknown "killer application" will determine many of the challenges and the solutions we must develop to make this technology work in practice.

32. Grossman R L. The case for cloud computing[J]. IT professional, 2009, 11(2): 23-27.

Abstract: To understand clouds and cloud computing, we must first understand the two different types of clouds. The author distinguishes between clouds that provide on-demand computing instances and those that provide on-demand computing capacity. Cloud computing doesn't yet have a standard definition, but a good working description of it is to say that clouds, or clusters of distributed computers, provide on-demand resources and services over a network, usually the Internet, with the scale and reliability of a data center.

33. Hofmann P, Woods D. Cloud computing: the limits of public clouds for business applications[J]. Internet Computing, IEEE, 2010, 14(6): 90-93.

Abstract: The cloud computing model - especially the public cloud - is unsuited to many business applications and is likely to remain so for many years due to fundamental limitations in architecture and design. Enterprises that move their IT to the cloud are likely to encounter challenges such as security, interoperability, and limits on their ability to tailor their ERP to their business processes. The cloud can be a revolutionary technology, especially for small startups, but the benefits wane for larger enterprises with more complex IT needs. Utility computing still cannot match the "plug-and-play" simplicity of electricity. On the other hand, private clouds offer the benefits like scale and virtualization with fewer drawbacks.

34. **Wang L, Tao J, Kunze M, et al. Scientific cloud computing: Early definition and experience[C]. High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on. Ieee, 2008: 825-830.**

Abstract: Cloud computing emerges as a new computing paradigm which aims to provide reliable, customized and QoS guaranteed computing dynamic environments for end-users. This paper reviews recent advances of Cloud computing, identifies the concepts and characters of scientific Clouds, and finally presents an example of scientific Cloud for data centers

35. **Boyd D, Crawford K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon[J]. Information, Communication & Society, 2012, 15(5): 662-679.**

Abstract: The era of Big Data has begun. Computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are clamoring for access to the massive quantities of information produced by and about people, things, and their interactions. Diverse groups argue about the potential

benefits and costs of analyzing genetic sequences, social media interactions, health records, phone logs, government records, and other digital traces left by people. Significant questions emerge. Will large-scale search data help us create better tools, services, and public goods? Or will it usher in a new wave of privacy incursions and invasive marketing? Will data analytics help us understand online communities and political movements? Or will it be used to track protesters and suppress speech? Will it transform how we study human communication and culture, or narrow the palette of research options and alter what ćresearchô means? Given the rise of Big Data as a socio-technical phenomenon, we argue that it is necessary to critically interrogate its assumptions and biases. In this article, we offer six provocations to spark conversations about the issues of Big Data: a cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis, and mythology that provokes extensive utopian and dystopian rhetoric.

36. Bendat J S, Piersol A G. Random data analysis and measurement procedures[J]. Measurement Science and Technology, 2000, 11(12): 1825.

Abstract: This is a new edition of a book on random data analysis which has been on the market since 1966 and which was extensively revised in 1971. The book has been a bestseller since. It has been fully updated to cover new procedures developed in the last 15 years and extends the discussion to a broad range of applied fields, such as aerospace, automotive industries or biomedical research. The primary purpose of this book is to provide a practical reference and tool for working engineers and scientists investigating dynamic data or using statistical methods to solve engineering problems. It is comprehensive and self-contained and expands the coverage of the theory, including derivations of the key relationships in probability and random-process theory not usually found to such extent in a book of this kind. It could well be used as a teaching textbook for advanced courses on the analysis of random processes. The first four chapters present the background material on descriptions of data, properties of linear systems and statistical principles. They also include probability distribution

formulas for one-, two- and higher-order changes of variables. Chapter five gives a comprehensive discussion of stationary random-process theory, including material on wave-number spectra, level crossings and peak values of normally distributed random data. Chapters six and seven develop mathematical relationships for the detailed analysis of single input/output and multiple input/output linear systems including algorithms. In chapters eight and nine important practical formulas to determine statistical errors in estimates of random data parameters and linear system properties from measured data are derived. Chapter ten deals with data aquisition and processing, including data qualification. Chapter eleven describes methods of data analysis such as data preparation, Fourier transforms, probability density functions, auto- and cross-correlation, spectral functions, joint record functions and multiple input/output functions. Chapter twelve shows how to handle nonstationary data analysis, classification of nonstationary data, probability structure of nonstationary data, calculation of nonstationary mean values or mean square values, correlation structures of nonstationary data and spectral structures of nonstationary data. The last chapter deals with the Hilbert transform including applications for both nondispersive and dispersive propagation problems. All chapters include many illustrations and references as well as examples and problem sets. This allows the reader to use the book for private study purposes. Altogether the book can be recommended for practical working engineers and scientists to support their daily work, as well as for university readers as a teaching textbook in advanced courses.

37. Weir B S. Genetic data analysis. Methods for discrete population genetic data[M]. Sinauer Associates, Inc. Publishers, 1990.

Abstract: This book describes, in detail, statistical methods used in the analysis of population genetic data of a discrete (enumeration) nature, such as genotype frequencies. It is not concerned with the analysis of continuously variable traits. It is the author's hope that the book will bridge the gap between Elandt-Johnson's Probability Models and Statistical Methods in Genetics, published 20 years ago, and

the range of techniques currently encountered in the scientific literature. The introductory chapter deals with the nature of genetic data and sampling, notation and terminology, and discusses the historical controversy over the goodness of fit of Mendel's experimental data to theoretical expectation. There follow chapters on the following: estimating frequencies (multinomial genotypic counts, likelihood estimation, etc); disequilibrium (Hardy-Weinberg, linkage); measures of gene diversity and heterozygosity; population structure (including genetic distance); analyses between generations (including estimation of outcrossing, selection and linkage, and inferences about paternity); molecular data (restriction fragment analysis, DNA sequence comparisons, etc); phylogeny construction. There are appendices on statistical distributions, random numbers, and FORTRAN source listings for programs on linkage disequilibrium, F-statistics (Wright) and DNA sequence manipulation. Each chapter has a summary, and exercises for the reader to carry out.

38. Hollands R G. Will the real smart city please stand up? Intelligent, progressive or entrepreneurial?[J]. City, 2008, 12(3): 303-320.

Abstract: Debates about the future of urban development in many Western countries have been increasingly influenced by discussions of smart cities. Yet despite numerous examples of this ôurban labellingôphenomenon, we know surprisingly little about so - called smart cities, particularly in terms of what the label ideologically reveals as well as hides. Due to its lack of definitional precision, not to mention an underlying self-congratulatory tendency, the main thrust of this article is to provide a preliminary critical polemic against some of the more rhetorical aspects of smart cities. The primary focus is on the labelling process adopted by some designated smart cities, with a view to problematizing a range of elements that supposedly characterize this new urban form, as well as question some of the underlying assumptions/contradictions hidden within the concept. To aid this critique, the article explores to what extent labelled smart cities can be understood as a high - tech variation of the ôentrepreneurial cityô, as well as speculates on some general principles

which would make them more progressive and inclusive.

39. Karnouskos S, De Holanda T N. Simulation of a smart grid city with software agents[C]. Computer Modeling and Simulation, 2009. EMS'09. Third UKSim European Symposium on. IEEE, 2009: 424-429.

Abstract: In the future smart city, new information and communication technologies will enable a better management of the available resources. The future smart grid infrastructure is emerging as a complex system where fine-grained monitoring and control of energy generating and/or consuming entities within the electricity network is possible. This will result to better approaches that will boost energy efficiency. A simulation of a dynamic ecosystem such as the smart city, will enable us to test new concepts and resource-optimization approaches. Therefore we have analyzed, designed, and build a simulator based on software agents that attempts to create the dynamic behavior of a smart city. It simulates discrete heterogeneous devices that consume and/or produce energy, that are able to act autonomously and collaborate. The behavior of these devices and their groupings e.g. smart houses, has been modeled in order to map as near as possible the real behavior patterns of the respective physical objects.

40. Barrett M A, Humblet O, Hiatt R A, et al. Big Data and Disease Prevention: From Quantified Self to Quantified Communities[J]. Big Data, 2013.

Abstract: Big data is often discussed in the context of improving medical care, but it also has a less appreciated but equally important role to play in preventing disease. Big data can facilitate action on the modifiable risk factors that contribute to a large fraction of the chronic disease burden, such as physical activity, diet, tobacco use, and exposure to pollution. It can do so by facilitating the discovery of risk factors for disease at population, subpopulation, and individual levels, and by improving the effectiveness of interventions to help people achieve healthier behaviors in healthier

environments. In this article, we describe new sources of big data in population health, explore their applications, and present two case studies illustrating how big data can be leveraged for prevention. We also discuss the many implementation obstacles that must be overcome before this vision can become a reality.

41. Madden S. From databases to big data[J]. Internet Computing, IEEE, 2012, 16(3): 4-6.

Abstract: There is a tremendous amount of buzz around the concept of "big data." In this article, the author discusses the origins of this trend, the relationship between big data and traditional databases and data processing platforms, and some of the new challenges that big data presents.

42. Fan W. Querying big social data[M]. Big Data. Springer Berlin Heidelberg, 2013: 14-28.

Abstract: Big data poses new challenges to query answering, from computational complexity theory to query evaluation techniques. Several questions arise. What query classes can be considered tractable in the context of big data? How can we make query answering feasible on big data? What should we do about the quality of the data, the other side of big data? This paper aims to provide an overview of recent advances in tackling these questions, using social network analysis as an example.

43. Davenport T H, Barth P, Bean R. How 'Big Data'is Different[J]. MIT Sloan Management Review, 2012, 54(1): 22-24.

Abstract: These days, many people in the information technology world and in corporate boardrooms are talking about ñbig data.ò Many believe that, for companies that get it right, big data will be able to unleash new organizational capabilities and value.

44. **Begoli E, Horey J. Design Principles for Effective Knowledge Discovery from Big Data[C]. Software Architecture (WICSA) and European Conference on Software Architecture (ECSA), 2012 Joint Working IEEE/IFIP Conference on. IEEE, 2012: 215-218.**

Abstract: Big data phenomenon refers to the practice of collection and processing of very large data sets and associated systems and algorithms used to analyze these massive datasets. Architectures for big data usually range across multiple machines and clusters, and they commonly consist of multiple special purpose sub-systems. Coupled with the knowledge discovery process, big data movement offers many unique opportunities for organizations to benefit (with respect to new insights, business optimizations, etc.). However, due to the difficulty of analyzing such large datasets, big data presents unique systems engineering and architectural challenges. In this paper, we present three system design principles that can inform organizations on effective analytic and data collection processes, system organization, and data dissemination practices. The principles presented derive from our own research and development experiences with big data problems from various federal agencies, and we illustrate each principle with our own experiences and recommendations.

45. **Rabl T, Gómez-Villamor S, Sadoghi M, et al. Solving big data challenges for enterprise application performance management[J]. Proceedings of the VLDB Endowment, 2012, 5(12): 1724-1735.**

Abstract: As the complexity of enterprise systems increases, the need for monitoring and analyzing such systems also grows. A number of companies have built sophisticated monitoring tools that go far beyond simple resource utilization reports. For example, based on instrumentation and specialized APIs, it is now possible to monitor single method invocations and trace individual transactions across geographically distributed systems. This high-level of detail enables more precise

forms of analysis and prediction but comes at the price of high data rates (i.e., big data). To maximize the benefit of data monitoring, the data has to be stored for an extended period of time for ulterior analysis. This new wave of big data analytics imposes new challenges especially for the application performance monitoring systems. The monitoring data has to be stored in a system that can sustain the high data rates and at the same time enable an up-to-date view of the underlying infrastructure. With the advent of modern key-value stores, a variety of data storage systems have emerged that are built with a focus on scalability and high data rates as predominant in this monitoring use case. In this work, we present our experience and a comprehensive performance evaluation of six modern (open-source) data stores in the context of application performance monitoring as part of CA Technologies initiative. We evaluated these systems with data and workloads that can be found in application performance monitoring, as well as, on-line advertisement, power monitoring, and many other use cases. We present our insights not only as performance results but also as lessons learned and our experience relating to the setup and configuration complexity of these data stores in an industry setting.

46. LIN F, HU M, JIANG Y, et al. Architecture and Related Techniques of a Power Dispatching Data Platform[J]. Automation of Electric Power Systems, 2007, 1: 012.

Abstract: Based on a review of the present status of power dispatching application data, the reference architecture of a power dispatching data platform is proposed. The main techniques including data input/output, data model and storage, data exchange and sharing, data processing, data service and data presentation are presented. Some related techniques such as data standards, data warehouse and data mining, as well as the integration with DMIS are also discussed.

47. Yang F, Shanmugasundaram J, Yerneni R. A Scalable Data Platform for a Large Number of Small Applications[C]. CIDR. 2009, 1(3): 11.

Abstract: As a growing number of websites open up their APIs to external application developers (e.g., Facebook, Yahoo! Widgets, Google Gadgets), these websites are facing an intriguing scalability problem: while each user-generated application is by itself quite small (in terms of size and throughput requirements), there are many many such applications.

48. Zikopoulos P C, Eaton C, DeRoos D, et al. Understanding big data[J]. New York et al: McGraw-Hill, 2012.

Abstract: Big Data represents a new era in data exploration and utilization, and IBM is uniquely positioned to help clients navigate this transformation. This book reveals how IBM is leveraging open source Big Data technology, infused with IBM technologies, to deliver a robust, secure, highly available, enterprise-class Big Data platform. The three defining characteristics of Big Data-volume, variety, and velocity are discussed. You'll get a primer on Hadoop and how IBM is hardening it for the enterprise, and learn when to leverage IBM InfoSphere BigInsights (Big Data at rest) and IBM InfoSphere Streams (Big Data in motion) technologies. Industry use cases are also included in this practical guide. Learn how IBM hardens Hadoop for enterprise-class scalability and reliability. Gain insight into IBM's unique in-motion and at-rest Big Data analytics platform. Learn tips and tricks for Big Data use cases and solutions. Get a quick Hadoop primer.

49. Jiyuan W X F A N. Construction of Common Data Platform in the Power Dispatcher Center[J]. Automation of Electric Power Systems, 2006, 22: 022.

Abstract: Based on IEC 61970/IEC 61968 standards and real-time CORBA technologies,this paper presents a new solution to construct a common data platform in the power dispatcher centers step by step.Three aspects are studied,such as common data management,data access pattern,and interconnection among

systems.The goal is to achieve the centralized modeling and maintaining of power system common data,enforce the integrity,security and maintainability of common information,and reduce the complexity and cost of interconnecting in different automation systems,and different power dispatcher centers.

50. Keivanloo I, Forbes C, Hmood A, et al. A Linked Data platform for mining software repositories[C]. Mining Software Repositories (MSR), 2012 9th IEEE Working Conference on. IEEE, 2012: 32-35.

Abstract: The mining of software repositories involves the extraction of both basic and value-added information from existing software repositories. The repositories will be mined to extract facts by different stakeholders (e.g. researchers, managers) and for various purposes. To avoid unnecessary pre-processing and analysis steps, sharing and integration of both basic and value-added facts are needed. In this research, we introduce SeCold, an open and collaborative platform for sharing software datasets. SeCold provides the first online software ecosystem Linked Data platform that supports data extraction and on-the-fly inter-dataset integration from major version control, issue tracking, and quality evaluation systems. In its first release, the dataset contains about two billion facts, such as source code statements, software licenses, and code clones from 18 000 software projects. In its second release the SeCold project will contain additional facts mined from issue trackers and versioning systems. Our approach is based on the same fundamental principle as Wikipedia: researchers and tool developers share analysis results obtained from their tools by publishing them as part of the SeCold portal and therefore make them an integrated part of the global knowledge domain. The SeCold project is an official member of the Linked Data dataset cloud and is currently the eighth largest online dataset available on the Web.

51. Foster I, Vockler J, Wilde M, et al. Chimera: A virtual data system for representing, querying, and automating data derivation[C]. Scientific and

Statistical Database Management, 2002. Proceedings. 14th International Conference on. IEEE, 2002: 37-46.

Abstract: A lot of scientific data is not obtained from measurements but rather derived from other data by the application of computational procedures. We hypothesize that explicit representation of these procedures can enable documentation of data provenance, discovery of available methods, and on-demand data generation (so-called "virtual data"). To explore this idea, we have developed the Chimera virtual data system, which combines a virtual data catalog for representing data derivation procedures and derived data, with a virtual data language interpreter that translates user requests into data definition and query operations on the database. We couple the Chimera system with distributed "data grid" services to enable on-demand execution of computation schedules constructed from database queries. We have applied this system to two challenge problems, the reconstruction of simulated collision event data from a high-energy physics experiment, and searching digital sky survey data for galactic clusters, with promising results.

52. Bunch C, Chohan N, Krintz C, et al. An evaluation of distributed datastores using the AppScale cloud platform[C]. Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on. IEEE, 2010: 305-312.

Abstract: We present new cloud support that employs a single API the Datastore API from Google App Engine (GAE) to interface to different open source distributed database technologies. We employ this support to "plug in" these technologies to the API so that they can be used by web applications and services without modification. The system facilitates an empirical evaluation and comparison of these disparate systems by web software developers, and reduces the barrier to entry for their use by automating their configuration and deployment.

53. **Furht B. Cloud computing fundamentals[M]. Handbook of cloud computing. Springer US, 2010: 3-19.**

Abstract: In the introductory chapter we define the concept of cloud computing and cloud services, and we introduce layers and types of cloud computing. We discuss the differences between cloud computing and cloud services. New technologies that enabled cloud computing are presented next. We also discuss cloud computing features, standards, and security issues. We introduce the key cloud computing platforms, their vendors, and their offerings. We discuss cloud computing challenges and the future of cloud computing.

54. **Katzan H. Cloud computing, I-Service, and IT service provisioning[J]. Journal of Service Science, 2011, 1(2): 57-64.**

Abstract: Cloud computing is an architecture for providing computing service via the Internet. Use of the term ñcloudò is a metaphor for the representation of the Internet used in most systems diagrams. In this case, the Internet is the transport mechanism between a client and a server located somewhere in cyberspace, as compared to having computer applications residing on an ñon premisesò computer. Adoption of cloud computing practically eliminates two ongoing problems in IT service provisioning: the upfront costs of acquiring computational resources and the time delay of building and deploying software applications. This paper covers both subjects.

55. **Chow R, Golle P, Jakobsson M, et al. Controlling data in the cloud: outsourcing computation without outsourcing control[C]. Proceedings of the 2009 ACM workshop on Cloud computing security. ACM, 2009: 85-90.**

Abstract: Cloud computing is clearly one of today's most enticing technology areas due, at least in part, to its cost-efficiency and flexibility. However, despite the surge in

activity and interest, there are significant, persistent concerns about cloud computing that are impeding momentum and will eventually compromise the vision of cloud computing as a new IT procurement model. In this paper, we characterize the problems and their impact on adoption. In addition, and equally importantly, we describe how the combination of existing research thrusts has the potential to alleviate many of the concerns impeding adoption. In particular, we argue that with continued research advances in trusted computing and computation-supporting encryption, life in the cloud can be advantageous from a business intelligence standpoint over the isolated alternative that is more common today.

56. Cheng J, Bell D A, Liu W. An algorithm for Bayesian belief network construction from data[C]. proceedings of AI & STAT'97. 1997: 83-90.

Abstract: This paper presents an efficient algorithm for constructing Bayesian belief networks from databases. The algorithm takes a database and an attributes ordering (i.e., the causal attributes of an attribute should appear earlier in the order) as input and constructs a belief network structure as output.

57. Sultan N. Cloud computing for education: A new dawn?[J]. International Journal of Information Management, 2010, 30(2): 109-116.

Abstract: Educational establishments continue to seek opportunities to rationalize the way they manage their resources. The economic crisis that befell the world following the near collapse of the global financial system and the subsequent bailouts of local banks with billions of tax payersô money will continue to affect educational establishments that are likely to discover that governments will have less money than before to invest in them. It is argued in this article that cloud computing is likely to be one of those opportunities sought by the cash-strapped educational establishments in these difficult times and could prove to be of immense benefit (and empowering in some situations) to them due to its flexibility and pay-as-you-go cost structure. Cloud

computing is an emerging new computing paradigm for delivering computing services. This computing approach relies on a number of existing technologies, e.g., the Internet, virtualization, grid computing, Web services, etc. The provision of this service in a pay-as-you-go way through (largely) the popular medium of the Internet gives this service a new distinctiveness. In this article, some aspects of this distinctiveness will be highlighted and some light will be shed on the current concerns that might be preventing some organizations from adopting it.

58. Marston S, Li Z, Bandyopadhyay S, et al. Cloud computing—The business perspective[J]. Decision Support Systems, 2011, 51(1): 176-189.

Abstract: The evolution of cloud computing over the past few years is potentially one of the major advances in the history of computing. However, if cloud computing is to achieve its potential, there needs to be a clear understanding of the various issues involved, both from the perspectives of the providers and the consumers of the technology. While a lot of research is currently taking place in the technology itself, there is an equally urgent need for understanding the business-related issues surrounding cloud computing. In this article, we identify the strengths, weaknesses, opportunities and threats for the cloud computing industry. We then identify the various issues that will affect the different stakeholders of cloud computing. We also issue a set of recommendations for the practitioners who will provide and manage this technology. For IS researchers, we outline the different areas of research that need attention so that we are in a position to advice the industry in the years to come. Finally, we outline some of the key issues facing governmental agencies who, due to the unique nature of the technology, will have to become intimately involved in the regulation of cloud computing.

59. Buyya R, Yeo C S, Venugopal S, et al. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility[J]. Future Generation computer systems, 2009, 25(6): 599-616.

Abstract: With the significant advances in Information and Communications Technology (ICT) over the last half century, there is an increasingly perceived vision that computing will one day be the 5th utility (after water, electricity, gas, and telephony). This computing utility, like all other four existing utilities, will provide the basic level of computing service that is considered essential to meet the everyday needs of the general community. To deliver this vision, a number of computing paradigms have been proposed, of which the latest one is known as Cloud computing. Hence, in this paper, we define Cloud computing and provide the architecture for creating Clouds with market-oriented resource allocation by leveraging technologies such as Virtual Machines (VMs). We also provide insights on market-based resource management strategies that encompass both customer-driven service management and computational risk management to sustain Service Level Agreement (SLA)-oriented resource allocation. In addition, we reveal our early thoughts on interconnecting Clouds for dynamically creating global Cloud exchanges and markets. Then, we present some representative Cloud platforms, especially those developed in industries, along with our current work towards realizing market-oriented resource allocation of Clouds as realized in Aneka enterprise Cloud technology. Furthermore, we highlight the difference between High Performance Computing (HPC) workload and Internet-based services workload. We also describe a meta-negotiation infrastructure to establish global Cloud exchanges and markets, and illustrate a case study of harnessing ‘Storage Clouds’ for high performance content delivery. Finally, we conclude with the need for convergence of competing IT paradigms to deliver our 21st century vision.

60. Shawish A, Salama M. Cloud Computing: Paradigms and Technologies[M]. Inter-cooperative Collective Intelligence: Techniques and Applications. Springer Berlin Heidelberg, 2014: 39-67.

Abstract: Cloud Computing has recently emerged as a compelling paradigm for

managing and delivering services over the internet. It is rapidly changing the landscape of information technology, and ultimately turning the long-held promise of utility computing into a reality. With such speedy progressing and emerging, it becomes crucial to understand all aspects about this technology. This chapter provides a comprehensive overview on the Cloud's anatomy, definition, characteristic, affects, architecture, and core technology. It clearly classifies the Cloud's deployment and service models, providing a full description of the Cloud services vendors. The chapter also addresses the customer-related aspects such as the Service Level Agreement, service cost, and security issues. Finally, it covers detailed comparisons between the Cloud Computing paradigm and other existing ones in addition to its significant challenges. By that, the chapter provides a complete overview on the Cloud Computing and paves the way for further research in this area.

61. Agrawal D, Das S, El Abbadi A. Big data and cloud computing: current state and future opportunities[C]. Proceedings of the 14th International Conference on Extending Database Technology. ACM, 2011: 530-533.

Abstract: Scalable database management systems (DBMS)---both for update intensive application workloads as well as decision support systems for descriptive and deep analytics---are a critical part of the cloud infrastructure and play an important role in ensuring the smooth transition of applications from the traditional enterprise infrastructures to next generation cloud infrastructures. Though scalable data management has been a vision for more than three decades and much research has focussed on large scale data management in traditional enterprise setting, cloud computing brings its own set of novel challenges that must be addressed to ensure the success of data management solutions in the cloud environment. This tutorial presents an organized picture of the challenges faced by application developers and DBMS designers in developing and deploying internet scale applications. Our background study encompasses both classes of systems: (i) for supporting update heavy

applications, and (ii) for ad-hoc analytics and decision support. We then focus on providing an in-depth analysis of systems for supporting update intensive web-applications and provide a survey of the state-of-the-art in this domain. We crystallize the design choices made by some successful systems large scale database management systems, analyze the application demands and access patterns, and enumerate the desiderata for a cloud-bound DBMS.

62. **Schadt E E, Linderman M D, Sorenson J, et al. Computational solutions to large-scale data management and analysis[J]. Nature Reviews Genetics, 2010, 11(9): 647-657.**

Abstract: Today we can generate hundreds of gigabases of DNA and RNA sequencing data in a week for less than US$5,000. The astonishing rate of data generation by these low-cost, high-throughput technologies in genomics is being matched by that of other technologies, such as real-time imaging and mass spectrometry-based flow cytometry. Success in the life sciences will depend on our ability to properly interpret the large-scale, high-dimensional data sets that are generated by these technologies, which in turn requires us to adopt advances in informatics. Here we discuss how we can master the different types of computational environments that exist ð such as cloud and heterogeneous computing ð to successfully tackle our big data problems.