

## 奥巴马政府推出大数据计划，宣布投资大数据领域

# 美国将发展大数据提升到战略层面

赛迪智库软件与信息服务研究所

2012年3月,美国奥巴马政府宣布推出“大数据的研究和发展计划”。该计划涉及美国国家科学基金、美国国家卫生研究院、美国能源部、美国国防部、美国国防部高级研究计划局、美国地质勘探局等6个联邦政府部门,承诺将投资两亿多美元,大力推动和改善与大数据相关的收集、组织和分析工具及技术,以推进从大量的、复杂的数据集中获取知识和洞见的能力。

美国奥巴马政府宣布投资大数据领域,是大数据从商业行为上升到国家战略的分水岭,表明大数据正式提升到战略层面,大数据在经济社会各个层面、各个领域都开始受到重视。

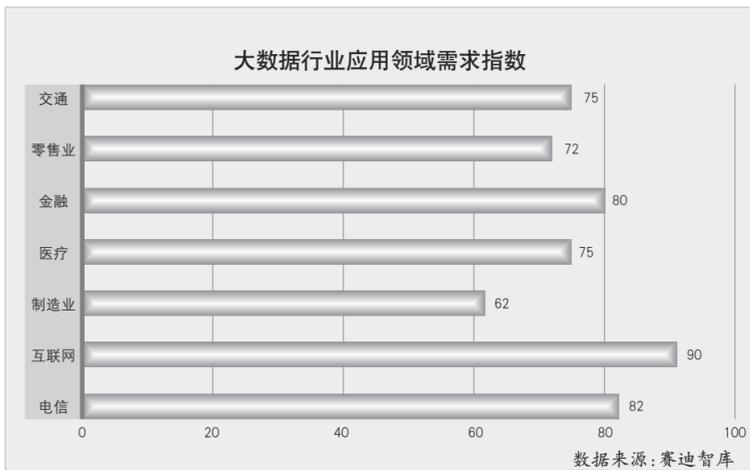
### 国家科学基金会：

#### 推进大数据科学核心技术

美国国家科学基金会和美国国家卫生研究院将对大数据进行联合招标,旨在提高提取、分析信息的能力。

美国国家科学基金会和美国国家卫生研究院将对大数据进行联合招标,旨在改进核心科学与技术手段,提高从各种大型数据集中提取重要信息并对其进行有效管理、分析和可视化的能力,加速科技成果的产生,并带领国家进入一些全新的、以往不可企及的研究领域。其中,卫生研究院对于与健康与疾病相关的数据集尤其感兴趣,包括影像、分子、细胞、电生理学、化学、行为、流行病学、临床医学数据集。国家科学基金会除为大数据招标提供资金维持其关注的基础研究外,还正在实施一个全面的、长期的战略,包括:开发新的方法,以便更有效地从数据中进行知识获取;相关基础设施投资,用于大数据科研团体的管理、组织和数据提供等;研究新的教育和人才培养方法。

采取的具体举措有:鼓励研究型大学设立跨学科的研究生专业课程,以培养新一代数据科学家和工程师人才;向加利福尼亚大学伯克利分校的计算开发项目投资1000万美元,此项目计划集成3种强大的数据转信息方法,包括机器学习、云计算和众包;为“EarthCube”提供第一阶段资金支持,该系统将允许地球学家获取、分析和共享与地球相关的信息;向一个研究培训小组发放200万美元的奖金,用于支持一项大学生培



训计划,教授他们如何利用图形和可视化工具解析复杂数据;为一个由统计学家和生物学家组成的专业研究团体提供140万美元的研发资金,资助他们研究蛋白质结构和生物途径;召集各个学科和领域的研究人员,共同探讨如何利用大数据转变教育与学习模式。

### 国防部：

#### 利用数据支持决策

美国国防部每年将投资2.5亿美元支持研究计划,旨在创新使用海量数据的方式。

美国国防部“在大数据上压下了巨大赌注”,他们每年将投资2.5亿美元左右(其中6000万美元用于支持新的研究项目),在各个军事部门开展一系列研究计划,旨在以创新方式使用海量数据,通过感知、认知和决策支持的结合,建立真正的、能够独立完成操控并做出决策的自治式系统;提高作战人员和分析师的环境与状况感知能力,增强对任务和流程的支持。国防部的目标是将分析师从任意语言文字资料中提取信息的能力提高100倍,同时希望他们观察到的目标、活动和事件的数量也获得相同幅度的提升。

开展的具体项目包括:多尺度异常检测项目,旨在解决大规模数据集的异常检测和特征化。目前多尺度异常检测应用程序能够进行内部威胁检测,以及在日常网络活动环

境中检测单独的异常行动。而洞悉计划则主要解决现有情报、监视和侦察系统的不足,进行自动化和人机集成推理,使得能够提前对时间敏感的更大潜在威胁进行分析。该计划旨在开发出资源管理系统,通过分析图像和非图像的传感器信息和其他来源的信息,进行网络威胁的自动识别和非常规的战争行为等。

此外,美国国防部高级研究计划局将开展XDATA项目。该项目计划在未来4年每年投入2500万美元开发能够分析海量半结构化数据和非结构化数据的计算技术和软件工具。需要解决的核心问题包括:开发可扩展的算法,用于处理分布式数据存储库中的不规则数据;创建有效的人机交互工具,用于支持面向各种处理任务的快速定制视觉分析。XDATA项目将支持开放源软件工具集,以帮助开发机构灵活开发软件,使用户能够尽快实现海量数据处理能力,与特定国防应用的任务数据流保持同步。

### 国家卫生研究院：

#### 免费开放千人基因组计划数据

免费开放由国际千人基因组计划创建的人类遗传变异研究数据集,供研究人员自由访问和使用。

美国国家卫生研究院宣布,由国际千

人基因组计划创建的最大的人类遗传变异研究数据集在亚马逊网站云服务(AWS)上免费公开。截至目前,数据量已经达到大约200TB,相当于1600万个塞满文本资料的文件柜或3万多个标准DVD的容量。该数据集的规模极为庞大,几乎没有哪个研究机构具有足够的计算能力对其中进行有效利用。现在AWS将千人基因组计划数据集免费公开,供研究人员自由访问和使用,他们只需为自己使用的计算服务付费。

### 能源部：

#### 通过高级计算技术加速科学发现

能源部将斥资2500万美元建立可扩展数据管理与可视化研究所,帮助科学家对数据进行有效管理。

美国能源部将斥资2500万美元建立可扩展数据管理、分析与可视化(SDAV)研究所。在能源部劳伦斯伯克利国家实验室(Lawrence Berkeley National Laboratory)的组织下,SDAV研究所将汇集6大国家实验室和7所著名大学的专业知识和经验来开发新工具,帮助科学家对能源部超级计算机上的数据进行有效管理和可视化处理。此举将进一步简化和加速开发流程,使科学家们能够利用能源部的研究设施开展更加卓有成效的科学研究和发现。目前能源部的超级计算机上同时运行的数据流在规模和复杂性方面均呈现不断增长的势头,因此对这些新型工具的需求也更加迫切。开展的主要项目计划包括:

高性能存储系统能够对千万亿次的数据进行分析处理,从庞大的科学数据集提取信息,发现其主要特征并理解其间的关系。系统广泛适应于从能源部到电网,包括宇宙学和天气数据、传感器数据等。

生物和环境研究计划:大气辐射测量气候研究设施是一个多平台的科学用户设施,可提供重要的大气现象的精确观测研究。主要用于应对从数以百计的文件中迅速采集和提交解决方法的挑战,以满足用户的需求。

美国核数据计划是一个多方面努力,涉及7个国家实验室和两所大学的项目,提供跨越多个领域、核物理、编译和交叉检查,对所有原子核的重要性质的相关实验结果,维护和广泛使用的专用数据库。

获取大数据的源头,再配合其数据库云服务器 Exadata、商务智能云服务器 Exalytics 以及其他相关软件(如 NoSQL 数据库、R 分析等)组成一个解决方案来全面满足用户的大数据需求。

### EMC 以云计算为平台

#### 做大数据处理业务

EMC 是大数据最早的“布道者”之一。相比竞争对手更多关注大型机器和更大型的管理软件。EMC 认为,应该利用云计算开放式、分布式和集群技术来做大数据处理。从整体来看,EMC 的大数据定位是通过提供解决方案,使用户从数据源获得最大价值,增强灵活性并提高效率。EMC 大数据解决方案构架在弹性的、横向扩展存储平台上,运行特性设计的分析工具,可处理结构化和非结构化数据,在生产环境中 and 业务流程平台上提供可付诸实践的洞察力。

(赛迪智库软件与信息服务研究所)

## 国际观察

美国“大数据的研究和发展计划”的提出意味着美国政府将大数据的研发与应用提升到保障美国国家安全、加速科学研究步伐、引发教育和学习变革的高度,标志着大数据将从以往的商业行为上升到国家意志和国家战略。可以预见,国家数字主权将成为继边防、海防、空防之后,另一个大国博弈的空间。

刘琼

大数据计划是美国保持信息产业强国霸主地位的重要体现。

1993年9月,美国宣布“国家信息基础设施”计划,也称“信息高速公路”计划。从此,发展信息高速公路成为美国联邦政府的一项国策,成就了美国作为全球信息产业霸主的地位,使之得以一直引领着全球信息产业的发展方向。

随着社交网络、云计算、移动互联网的迅猛发展,数据规模呈爆发式增长,全球进入大数据时代。作为全球信息产业的的主导国,美国发布大数据战略既是其“国家信息基础设施”计划的延伸,更是其谋求保持新信息时代霸主地位的重要体现。

大数据战略是商业行为上升到国家科技战略的“分水岭”。

由于有望产生巨大经济效益,美国的麦肯锡、EMC、IBM 等企业提出大数据概念并加以推进,美国一些大型公司已经开始赞助大数据相关项目的竞赛,并且为高等院校的大数据研究提供资金。在这之后,大数据开始受到美国政府的重视。

各国以新一代信息产业为核心的新一轮产业革命正在兴起,目标是占领后危机时代经济发展的制高点。美国奥巴马政府出于对科技发展成果和前景的判断,认识到大数据对于未来国家发展的重要意义,将其定义为“未来的新石油”,赋予发展大数据非同一般的战略意义,将大数据升级为一项借助信息技术提升国家竞争力的宏观战略。

大数据战略是美国提高情报信息处理能力的需要。

美国的情报体系非常庞大,拥有16个机构、20万多名员工。即使如此庞大的体系,也越来越难以及时处理无人机、侦查机器人等各种平台传送回来的海量信息。美国反恐中心一天需要查看的数据库超过120个,每天处理1万至1.2万条情报信息。如果说,2001年“9·11”恐怖袭击事件前,美国情报部门的问题是没能共享信息,但改革后却陷入“信息爆炸”、难以处理的窘境。整合情报网络体系、提高情报信息处理能力的需求日益迫切。2011年12月25日美国底特律发生的未遂炸机事件将这一需求推升至空前的高度。美国总统奥巴马在2012年1月7日的讲话中要求情报部门必须“更快、更广”地分发报告,并加强情报分析和整合能力。大数据计划在一定程度上是对美国总统奥巴马上述指示的落实。

然而,大数据的应用是集成应用,需要集成人工智能、商业智能、数学算法、自然语言理解、信息技术等多个跨学科领域的技术成果,面临较高的技术和管理风险。且大数据的应用场景不同,需要采用的技术和方法也不同,是一对一的应用。因此,美国“大数据”战略能否如其所愿尚待实践检验。事实上,“大数据研究与开发计划”的一些项目已出现“可能实施不了”的声音。所以,大数据能否落地,尚需拭目以待。

## 传统 IT 巨头打造大数据综合解决方案

IBM、EMC、惠普等传统 IT 巨头凭借自身在数据库、商业智能、存储、数据分析等原本被软件产品独占的领域,通过并购、投资等方式,整合软硬件业务,推出面向大数据的一体化产品和解决方案,进而巩固在 IT 领域的龙头地位。

### IBM 以数据分析为核心

#### 打造综合集成能力

作为全球最重要的 IT 厂商,IBM 在大数据方面拥有3方面的独特优势。一是 IBM 拥有广泛且整合的信息与分析产品和服务组合,包括企业级的大数据平台、分析和决策解决方案。二是拥有运用经验的能力。在分析领域,IBM 已经成功实施过两千多个项目、9000多个相关的专业顾问以及专门的分析解决方案

◎国际传真

### 西班牙互联网

#### 流量将飞速增长

近日由思科公司发布的报告显示,预计至2016年,西班牙的 IP 流量将达到2011年水平的13倍,即平均每年达到3.8EB,每月达到320PB。基于这一预测,2011~2016年期间西班牙 IP 流量的年均增长率将高达67%。

报告中指出,至2016年,西班牙的联网设备量将可能达到2.58亿部,固定线路网络用户将达到3700万,而移动网络用户将达到660万。同时,宽带网络平均速度也将达到2011年水平的4倍,即从10.2Mbps升至40Mbps。值得一提的是,仅在2010年至2011年期间,西班牙的宽带网络平均速度就增长了57%,从6.5Mbps一跃至了10.2Mbps。

思科公司预计,截至2016年全球互联网用户数量将达到34亿,即占该年全球预估人口数量的45%左右。与这一预测水平相比,

中心,三是拥有先进的分析能力。IBM 拥有全球商业企业中最大的数学部门,每年有600项以上与分析相关的专利,有 IBM 沃生这种有学习能力的智慧分析系统。

### 惠普以大并购为途径

#### 实现价值链全覆盖

围绕大数据进行大并购,使惠普逐渐实现了大数据价值链的全覆盖,成为大数据一体化解决方案供应商。

2010年9月3日,惠普以23.5亿美元收购了存储企业3PAR,其平台成为增长最快的高端存储平台,同时也是惠普目前营收最大的存储产品阵列。2011年2月,惠普收购分析及数据管理软件公司 Vertica,2011年 Vertica 实现大数据收入

2011年全球互联网用户数量为20亿人,占全球总人口比例的28%。同样,预计至2016年全球网络视频时长将达到120万分钟,即约为833天。报告还预计,届时全球将有超过一半的互联网流量是通过无线连接实现的。

### 古巴将大力发展

#### 可再生能源

古巴政府日前表示,将在未来8年内实现国内可再生能源产量增加12%的目标。

为实现这一目标,古巴政府将着力加大对森林生物质发电及甘蔗能发电新技术的开发,同时加大力度推进对太阳能、风能及水资源等清洁能源的应用。

预计未来制糖工业将成为推动古巴可再生能源产业发展的主要支柱之一。在太阳能应用方面,预计2012年12月古巴将建立一个峰值功率达1MWp的光伏园区,2013年

8400万美元。2011年8月,惠普以100亿美元收购英国第二大软件厂商 Autonomy,其软件被设计用来识别结构化数据和非结构化数据之间的关系。

惠普将原有的技术能力和3PAR、Autonomy、Vertica的技术能力进行了融合创新,推出了面向大数据应用的系列解决方案。在开源机构 Wikibon 发布的2012年全球科技企业在大数据市场的营收排名榜上,惠普以5.5亿美元的规模位居第三位。

### Oracle 以数据库为中心

#### 全面满足用户需求

在大数据市场,Oracle 沿袭了其一贯的集成一体设计的思路,以大数据机作为

还将继续加大建设,实现总的峰值功率达到10MWp。Cubadebate 网站透露,事实上古巴全国已经确认的太阳能装机容量可以超过2000MW,但目前却仅有一些小型的光伏电站实现了与国家电力系统的联网并提供相关服务。

另据统计数字显示,2011年,古巴全国应用水力和风力资源发电的总量共替代了约合31150吨传统燃料的发电产量,成功减少了高达10万吨二氧化碳排放量,较1990年碳排放量减少了近20%。

### 南非长途光纤网

#### 第一阶段即将完工

7月1日,南非跨国电信集团 MTN 宣称,横跨南非主要城市中心的5000公里国家长途光纤网络即将完成第一阶段任务。

MTN 首席技术官 Kanagaratnam

Lambotharan 说,在杰斯顿和德班之间挖掘的第一条路线,占据了总路线的90%以上,利用浮动技术的光纤安装工作已经完成了大约50%的进度。

此外,两个主要的 MTN 节点,即 New Germany(南非的一个城市)节点和德班节点,在 MTN 的网络上已经开始启用。

由 Neotel 与 Vodacom 联合开发的这个项目,经过了精心设计和优化,用来连接主要的金融中心和经济枢纽,以及与国际海缆登陆点互连连接。

Kanagaratnam Lambotharan 说:“该网络被设计用来与 MTN 的国际电缆资产直接进行连接,从而能够与接入第一层互联网骨干网,以及与 MTN 在非洲和中东的其他业务进行高效率和高品质的连接。”

Kanagaratnam Lambotharan 补充说,通过网络可用性和应变能力的增强,将大大提高网络质量,可满足针对固定电话融合服务和带宽密集型内容日益增长的需求。(本栏目文章由工业和信息化部国际经济技术合作中心白雪曼、李雅提供)