大数据安全标准化白皮书

（2017）

全国信息安全标准化技术委员会
大数据安全标准特别工作组

2017年4月

大数据安全标准化白皮书

（2017）

全国信息安全标准化技术委员会
大数据安全标准特别工作组

2017年4月

编写单位

中国电子技术标准化研究院
清华大学
四川大学
阿里云计算有限公司
中电长城网际系统应用有限公司
阿里巴巴（北京）软件服务有限公司
国家信息安全工程技术研究中心
华为技术有限公司
中国电子科技网络信息安全有限公司
深信服科技股份有限公司
西安未来国际信息股份有限公司
广州赛宝认证中心服务有限公司
北京奇虎科技有限公司
中国移动通信集团公司
大唐联诚信息系统技术有限公司
深圳市腾讯计算机系统有限公司
北京京东叁佰陆拾度电子商务有限公司
英特尔（中国）有限公司
国际商业机器（中国）有限公司
肯睿（上海）软件有限公司
微软(中国)有限公司
中国信息安全认证中心
百度在线网络技术（北京）有限公司
北京数字认证股份有限公司
国家信息中心
北京赛博兴安科技有限公司
北京工业大学
医渡云（北京）技术有限公司

编写人员

王建民	陈兴蜀	刘贤刚	范科峰	叶润国	金　涛	陈雪秀	闵京华
江为强	谢安明	叶晓俊	韩晓露	贾　科	任兰芳	鲍旭华	吕　欣
陈海锋	程广明	李克鹏	路　琨	程海旭	望娅露	许怡娴	葛　龙
胡　影	王　超	刘　焱	梅婧婷	张　滨	冯运波	张　凡	符海芳
刘伯仲	蔡　磊	王惠莅	许东阳	黄少青	姚相振	周斌琦	朱红儒
葛小宇	曾志峰	王劲松	郑　斌	翟胜军	任泽君	李建伟	刘小茵
吴　迪	张　群	杨　震	张　峰	罗永刚	糜　君	阮树骅	龚洁中
何延哲	周　毅	刘　钊	彭　锡	宋玲娓	朱　悦	黄　敏	贾雪飞

版权声明：如需转载或引用，请注明出处。

目录 CONTENTS

第一章 导论 …… 1

1.1 背景 …… 1

1.2 目的及意义 …… 2

第二章 大数据安全法规政策和标准化现状 …… 3

2.1 大数据安全法律法规和政策 …… 3

2.1.1 国外数据安全法律法规和政策 …… 3

2.1.2 国内数据安全法律法规和政策 …… 7

2.2 主要标准化组织大数据安全工作情况 …… 9

2.2.1 ISO/IEC JTC1 …… 9

2.2.2 ITU–T …… 10

2.2.3 NIST …… 11

2.2.4 TC28 …… 11

2.2.5 TC260 …… 12

2.3 大数据安全相关标准现状 …… 12

2.3.1 传统数据安全标准规范 …… 13

2.3.2 个人信息安全标准规范 …… 15

2.3.3 大数据安全标准规范 …… 17

第三章 大数据安全标准体系 …… 21

3.1 大数据安全挑战 …… 21

3.1.1 技术平台角度 …… 21

3.1.2 数据应用角度 …… 22

3.2 大数据安全标准化需求 …… 24

目录 CONTENTS

3.3 大数据安全标准体系框架 …… 25
3.4 大数据安全标准规划 …… 27

第四章 大数据安全标准化工作建议 …… 31
4.1 加强大数据安全核心技术研究 …… 31
4.2 加快制定个人信息安全相关标准 …… 31
4.3 加快制定数据共享相关安全标准 …… 32
4.4 加快制定数据出境安全相关标准 …… 32
4.5 加快大数据安全审查支撑性标准研制 …… 33
4.6 大力推广大数据安全标准应用 …… 33
4.7 加强大数据安全标准化人才的培养 …… 33
4.8 深度参与大数据安全国际标准制定 …… 34

缩略语 …… 35

附录A 典型行业大数据应用和安全风险 …… 37
1. 安全大数据 …… 37
2. 电子政务大数据 …… 40
3. 健康医疗大数据 …… 43
4. 电商行业大数据 …… 44
5. 电信行业大数据 …… 46

目录 CONTENTS

附录B 大数据应用安全实践 …… 49
1. 阿里云大数据安全实践 …… 49
2. 百度大数据安全实践 …… 52
3. 华为大数据安全实践 …… 54
4. 京东大数据安全实践 …… 56
5. 奇虎360大数据安全实践 …… 58
6. 腾讯大数据安全实践 …… 59
7. 中国移动大数据安全实践 …… 61
8. Cloudera大数据安全实践 …… 64
9. Hadoop大数据安全实践 …… 66
10. IBM 大数据安全实践 …… 68
11. Microsoft大数据安全实践 …… 71

第一章 导论

1.1 背景

随着社会信息化和网络化的发展，数据爆炸式增长，大数据时代已经到来。大数据被誉为是“21世纪的钻石矿”，是国家基础性战略资源，正日益对国家治理能力、经济运行机制、社会生活方式以及各领域的生产、流通、分配、消费活动产生重要影响，各国政府都在积极推动大数据应用与发展。

大数据时代是机遇与挑战并存的时代。在大数据应用推广过程中，必须坚持安全与发展并重的方针，为大数据发展构建安全保障体系，在充分发挥大数据价值的同时，解决面临的数据安全和个人信息保护问题。大数据安全标准是大数据安全保障体系的重要组成部分，对其实施起到引领和指导性作用。为此，亟待从技术和产业发展角度加快推进大数据安全标准化工作，为我国大数据产业的健康发展提供有效支撑。

党中央、国务院高度重视大数据安全及其标准化工作，将其作为国家发展战略予以推动。2015年9月，国务院发布《促进大数据发展行动纲要》，要求“完善法规制度和标准体系”和“推进大数据产业标准体系建设”。2016年11月，第十二届全国人民代表大会常务委员会通过了《中华人民共和国网络安全法》，鼓励开发网络数据安全保护和利用技术。2016年12月，国家互联网信息办公室发布《国家网络空间安全战略》，在夯实网络安全基础的战略任务中，提出实施国家大数据战略、建立大数据安全管理制度、支持大数据信息技术创新和应用要求。全国人大常委会和工信部、公安部等部门为加快构建大数据安全保障体系，相继出台了《加强网络信息保护的决定》、《电信和互联网用户个人信息保护规定》等法规和

部门规章制度。与此同时，还发布了国家和行业的网络个人信息保护相关标准，开展了以数据安全为重点的网络安全防护检查。

为推动大数据安全标准化工作，全国信息安全标准化技术委员会（以下简称“全国信安标委”，委员会编号为TC260）下设的大数据安全标准特别工作组（SWG-BDS）启动了《大数据安全标准化白皮书》的编制工作。

1.2 目的及意义

本白皮书从法规、政策、标准和应用等角度，勾画出大数据安全的整体轮廓，综合分析大数据安全标准化需求，为我国后续的大数据安全标准化工作提供指导。

本白皮书介绍了国内外的大数据安全法规政策和标准化组织标准化工作现状，分析了大数据安全所面临的安全风险和挑战，制定了大数据安全标准化体系框架和近几年大数据安全标准工作规划，并提出了开展大数据安全标准化工作的建议。

本白皮书旨在全面、客观反映国内外大数据安全标准化相关工作基础和进展，根据业界最佳实践、认知水平，分享大数据安全标准特别工作组在大数据安全标准化领域的研究成果和实践经验，呼吁社会各界共同关注大数据安全的政策研究、技术投入和标准建设，为大数据产业的健康、安全、有序发展奠定坚实基础。

第二章 大数据安全法规政策和标准化现状

2.1 大数据安全法律法规和政策

随着各国对大数据安全重要性认识的不断加深，包括美国、英国、澳大利亚、欧盟和我国在内的很多国家和组织都制定了大数据安全相关的法律法规和政策来推动大数据利用和安全保护，在政府数据开放、数据跨境流通和个人信息保护等方向进行了探索与实践。

2.1.1 国外数据安全法律法规和政策

（一）政府数据开放相关法规和政策

政府数据开放是指在确保国家安全条件下，政府向公众开放财政、资源、人口等公共数据信息，以增强公众参与社会管理的意愿和能力，进而提升政府治理水平。美国将信息技术、数字战略、信息管理与开放政府治理有机结合，以数据开放作为新时期政府治理改革的突破口。美国为推动政府数据开放，发布的法规政策主要包括：

1）美国于1966年通过《信息自由法》，规定民众在获得行政信息方面的权利和行政机关在向民众提供行政信息方面的义务。该法秉持“以公开为原则、不公开为例外”的原则，规定政府有义务对公众的信息公开请求作出回应。

2）《开放政府指令》是美国政府2009年在数据开放方面的最新行动，确立了透明、参与和协作的开放政府三原则，要求政府通过网站发布数据等方式，使公众了解更多的政府信息，促进公共对话，提升公众对政府的信任感。

3）美国在2012 年5 月出台了《数字政府战略》，将政府开放数据作为电子政府发展的重要支撑，它要求政府数据在默认状态处于“开放和机器可读”，从而使得公众可随时随地访问高质量的政府数据信息和服务。

英国作为最早实施数据开放的国家，通过开放政府数据提升政府治理水平，并致力于将数据开放推广应用到公共服务、经济增长、反对腐败、加强民主等诸多方面。英国于2000年正式通过《信息自由法》，规定了任何人都有获取政府信息的权利，政府有答复公众请求的义务，同时给出了25种公开豁免情况。英国政府在2013年4月发布的《开放政府伙伴2013—2015英国国家行动方案》中进一步拓展数据开放承诺，表示将开放政府数据，以改善公共服务，促进经济增长、提高政府透明度。

欧盟十分重视政府数据开放，欧盟执行委员会承认政府部门资料开放使用的重要性，积极鼓励各成员国展开政府开放数据的行动。2005年出台的《欧洲透明度倡议》(ETI)是欧盟推行开放数据战略的基础，也是欧盟开放、透明、治理改革理念的继续与发展。《欧洲透明度倡议》目的在于建立信息再利用，包括监管公共部门的共同法律框架，消除公共信息垄断和不透明障碍。2010年11月，欧盟通信委员会向欧洲议会提交了《开放数据:创新、增长和透明治理的引擎》报告，该报告以开放数据为核心，制定了应对大数据安全挑战的战略。2011年12月12日，欧盟数字议程正式推进数据开放战略，将其作为实现欧盟“2020目标”的新路径与新动力。

（二）数据跨境流动相关法规和政策

当前，部分国家和地区在规范跨境转移个人数据的法律法规，以便对跨境数据接收地的法律环境提出要求，要求接收地法律能够提供与本国、本地个人数据保护法律相当的保护。规范个人数据跨境转移的根本目的，不是禁止个人数据跨境转移，而是要根据实际情况，确保本国、本地区公民数据在境外受到合理保护。总体上，基本立场是鼓励数据跨境自由流动。具体与数据跨境流动相关的政策及法规包括:

欧盟在1995年《数据保护指令》中确立了数据跨境传输的基本原则，对世界其他国家和地区的信息保护、信息流动产生了深远的影响。2015年通过的欧盟《通用数据保护条例》（GDPR）提出了更为苛刻的数据保护规定。GDPR的适用范围相比《数据保护指令》有所扩展，对于那些即使成立地在欧盟以外的机构来说，只要其在提供产品或者服务的过程中涉及欧盟境内个体的个人数据，将同样适用。在跨境数据流动方面，GDPR增加了新的制度安排，包括认证机制和行为准则等。

澳大利亚《隐私保护原则》第8条规定，数据主体获得明确告知后同意的，可以将个人数据传输至境外数据接收者。告知必须解释跨境传输带来的数据主体应当知道的后果或风险，包括:1)海外接收者可能不承担类似澳大利亚《隐私保护原则》规定的隐私保护义务；2)数据主体可能无法在海外司法管辖区获得救济；3)海外数据接收者可能会根据法律将个人信息披露给第三方，比如外国政府机构。

巴西司法部于2015年公布的《个人数据保护法》（草案）也借鉴了欧盟有关充分性数据保护的规定，要求跨境数据传输时，数据接收国的个人数据保护必须达到充分性保护的水平。

韩国2011年发布的《个人信息保护法》规定，如果个人信息的国际数据传输涉及第三方，那么必须取得用户的明确同意。2012年发布的《促进信息技术网络利用和信息保护法》要求则更为明确：如果用户的个人信息被转移到境外实体，在线服务提供商必须告知并获得用户的明示同意。

日本2015年修订的《个人信息保护法》规定，个人信息可以“传输到为日本个人信息保护委员会（PIPC）所认可的、与日本国内个人信息保护水平相当的国家或地区”。

亚太经合组织（APEC）于2003年发布了《APEC隐私保护框架》，它采取使用企业自律性隐私政策来保护跨境数据流动中数据（隐私）权利的做法，在《APEC隐私保护框架》下建立跨境隐私保护规则体系。APEC不

仅仅是将跨境隐私保护规则看作是跨境数据转移中保护隐私的一种手段，更将其作为执行《APEC隐私保护框架》的重要机制。

（三）个人数据保护相关法规和政策

1）欧盟《通用数据保护条例》

欧盟颁布的《通用数据保护条例》（GDPR）为一个数据隐私权标准法规，它旨在取代1995年发布的《数据保护指令》。GDPR主要为了保护欧盟公民个人数据的隐私权，并对数据保护规则进行了改革和更新。如GDPR引入了新型的数据主体权利，包括“数据可携带权”和“被遗忘权”，为个人有效行使权利提供了坚实的法律保障。GDPR要点包括个人拥有管理自己个人数据的权利、数据泄露获得通知的权利以及“被遗忘权”。GDPR由两部分组成：通用数据保护条例，这“将让人们更好地控制其个人数据”；数据保护指令，对于警察和刑事司法领域，这“可确保数据受害人、证人和犯罪嫌疑人在刑事调查或执法行动中受到应有的保护。”这些明确的法规特别适合数字时代，能提供强有力的保护，同时在欧洲数字单一市场创造机会和鼓励创新，将让公民和企业都受益。

2）美国健康保险携带和责任法案

1996年，美国总统签署适用于提供健康保健的医疗组织和其它符合健康计划组织的健康保险携带和责任法案（HIPAA）。HIPAA目标是确保健康信息的安全性和隐私性，其主要内容包括隐私条例和安全条例。隐私条例保护所有由适用实体保存的可识别个体的受保护健康信息（PHI）。根据美国卫生和福利部的规定，PHI包括以下数据信息：与个人以往、目前或将来的身体（或精神）健康或状况有关的数据；个人接受健康保健服务的相关数据；个人以往、目前或将来接受健康保健服务的费用支付相关的数据。隐私条例的基本规定是企业只有在隐私条例允许范围内或者获得数据主体的个人书面同意之后才能够披露PHI。隐私条例还包含了一些通知规定和管理规定，保证企业保持数据记录行为的恰当性，以及确保个人

明确自己受HIPAA条例保护的权利。安全条例包含了电子受保护健康信息（ePHI）的安全保护，规定了企业在其所有需要处理ePHI数据的系统里必须具有的策略、流程和报告机制。HIPAA还规定了用于保护ePHI的保密性、完整性和可获得性的具体实施规定。这些规定包含管理防护、物理防护、技术防护、组织要求、企业策略和流程。

3）美国家庭教育权和隐私权法案

1974年，美国联邦法案家庭教育权和隐私权法案（FERPA）出台，用以保护学生的个人可识别信息（PII）的安全，适用于教育行业。FERPA规定，未满18岁的学生或符合条件的学生家长可以查看并申请修正学生的教育记录。该法案还规定，学校必须取得学生家长或符合条件的学生的书面许可才能披露学生的个人信息。

2.1.2 国内数据安全法律法规和政策

我国在积极推动大数据产业发展的过程中，非常关注大数据安全问题，近几年发布了一系列大数据产业发展和安全保护相关的法律法规和政策。

2012年12月，针对数据应用过程中的个人信息保护问题，第十一届全国人民代表大会常务委员会通过了《全国人大常委会关于加强网络信息保护的决定》，该决定要求，国家保护能够识别公民个人身份和涉及公民个人隐私的电子信息，网络服务提供者和其它企事业单位应当采取技术措施和其它必要措施，确保信息安全，防止在业务活动中收集的公民个人电子信息泄露、损毁、丢失。在发生或者可能发生信息泄露、损毁、丢失的情况时，应当立即采取补救措施。

2013年7月，工业和信息化部公布了《电信和互联网用户个人信息保护规定》。该规定是对全国人大常委会《关于加强网络信息保护的决定》的贯彻落实，进一步明确了电信业务经营者、互联网信息服务提供者收

集、使用用户个人信息的规则和信息安全保障措施要求。2014年3月，我国新的《消费者权益保护法》正式实施。该法明确了消费者享有个人信息依法得到保护的权利，同时要求经营者采取技术措施和其他必要措施，确保个人信息安全，防止消费者个人信息泄露、丢失。

2015年8月，国务院印发《促进大数据发展行动纲要》（以下简称“行动纲要”），提出加快建设数据强国和释放数据红利，并加快政府数据开放共享，以提升治理能力。同时，行动纲要提出网络空间数据主权保护是国家安全的重要组成部分，要求“强化安全保障、提高管理水平，促进健康发展”，并探索完善安全保密管理规范措施，切实保障数据安全。在大数据安全标准方面，行动纲要提出要进一步完善法规制度和标准体系，大力推进大数据产业标准体系建设。

2016年3月，第十二届全国人大四次会议表决通过了《关于国民经济和社会发展第十三个五年规划纲要》（以下简称“十三五规划纲要”）。十三五规划纲要提出实施国家大数据战略，全面实施促进大数据发展行动，同时要强化信息安全保障。该规划纲要提出加强数据资源安全保护，具体表现为要建立大数据安全管理制度、实行数据资源分类分级管理和保障安全高效可信应用。

2016年11月，全国人民代表大会常务委员会发布了《中华人民共和国网络安全法》（以下简称“网络安全法”）。网络安全法定义网络数据为通过网络收集、存储、传输、处理和产生的各种电子数据，并鼓励开发网络数据安全保护和利用技术，促进公共数据资源开放，推动技术创新和经济社会发展。关于网络数据安全保障方面，网络安全法规定，要求网络运营者采取数据分类、重要数据备份和加密等措施，防止网络数据被窃取或者篡改，加强对公民个人信息的保护，防止公民个人信息被非法获取、泄露或者非法使用，要求关键信息基础设施的运营者在境内存储公民个人信息等重要数据，网络数据确实需要跨境传输时，需要经过安全评估和审批。

2016年12月，国家互联网信息办公室发布《国家网络空间安全战略》，提出要实施国家大数据战略，建立大数据安全管理制度，支持大数据、云计算等新一代信息技术创新和应用，为保障国家网络安全夯实产业基础。

2.2 主要标准化组织大数据安全工作情况

目前，多个标准化组织正在开展大数据和大数据安全相关标准化工作，主要有国际标准化组织/国际电工委员会下的ISO/IEC JTC1 WG9（大数据工作组）、ISO/IEC JTC1 SC27（信息安全技术分委员会）、国际电信联盟电信标准化部门（ITU-T）、美国国家标准与技术研究院（NIST）等。国内正在开展大数据和大数据安全相关标准化工作的标准化组织，主要有全国信息技术标准化委员会（以下简称“全国信标委”，委员会编号为TC28）和全国信安标委（TC260）等。

2.2.1 ISO/IEC JTC1

ISO/IEC JTC1 SC27是在ISO和IEC信息技术联合委员会（ISO/IEC JTC1）下属安全技术分委员会，成立于1990年，其工作范围涵盖信息和ICT（信息与通信技术）保护的标准开发，包括安全与隐私保护方面的方法、技术和指南。目前下设五个工作组，分别为信息安全管理体系工作组（WG1）、密码技术与安全机制工作组（WG2）、安全评价、测试和规范工作组（WG3）、安全控制与服务工作组（WG4）和身份管理与隐私保护技术工作组（WG5）。各工作组负责各自工作范围内的多项标准开发，并根据需要设立相应的研究项目。

其中，WG5负责身份管理和隐私保护相关标准的研制和维护。WG5结合其工作范围和重点，开发了标准路线图，概括了WG5已有标准项目、新工作项目提案，以及将来WG5可能涉及到的标准化主题等内容。WG5

工作组负责制定的隐私保护方面标准包括已发布的ISO/IEC 29100:2011《信息技术　安全技术　隐私保护框架》、ISO/IEC 29101:2013《信息技术　安全技术　隐私保护体系结构框架》、ISO/IEC 29190:2015《信息技术　安全技术　隐私保护能力评估模型》、ISO/IEC 29191:2012《信息技术　安全技术　部分匿名、部分不可链接鉴别要求》和ISO/IEC 27018:2014《信息技术　安全技术　可识别个人信息（PII）处理者在公有云中保护PII的实践指南》，即将发布的ISO/IEC 29134《信息技术　安全技术　隐私影响评估指南》和ISO/IEC 29151《信息技术　安全技术　可识别个人信息（PII）保护实践指南》，以及正在工作草案阶段的ISO/IEC 29184《在线隐私通知和准许指南》、ISO/IEC 27550《隐私保护工程》和ISO/IEC 27551《对ISO/IEC 27001在隐私保护管理方面的增强要求》。

ISO/IEC JTC1 WG9是ISO/IEC JTC1于2014年11月成立的大数据工作组，目前正在开展ISO/IEC 20546《信息技术　大数据　概述和词汇》和ISO/IEC 20547《信息技术　大数据参考架构》两项国际标准编制。ISO/IEC 20547为多部分标准，包括ISO/IEC TR 20547-1《第1部分：框架和应用过程》、ISO/IEC TR 20547-2《第2部分：用例和衍生需求》、ISO/IEC 20547-3《第3部分：参考架构》、ISO/IEC 20547-4《第4部分：安全与隐私保护》、ISO/IEC TR 20547-5《第5部分：标准路线图》。

其中，ISO/IEC 20547-4《信息技术　大数据参考架构　第4部分：安全与隐私保护》标准编制项目根据ISO/IEC JTC1 JAG（JTC1咨询小组）2016年3月巴黎会议决定被转交给了ISO/IEC JTC1 SC27，现由SC27下属WG4和WG5共同负责，并任命中国专家担任项目编辑。

2.2.2 ITU-T

ITU-T在2013年11月发布了《大数据：今天巨大，明天平常》报告，并在其下属相关研究组开展了多项大数据和大数据安全相关的标准化工作。

ITU-T SG13（聚焦于IMT-2020、云计算和可信网络基础设施的未来网络研究组）负责制定的大数据相关标准包括：已发布的ITU Y.3600《大数据　基于云计算的要求和能力》，以及在编制中的《大数据　元数据框架和概念模型》、《大数据　数据集成概述和功能要求》、《大数据　数据溯源要求》、《大数据交换框架和要求》、《数据存储联合的要求和能力》、《大数据即服务的功能架构》、《大数据　数据保全概述和要求》、《大数据驱动联网要求》、《基于DPI的大数据驱动联网框架》和《应用于网络大数据语境下的深度包检测机制》等。

ITU-T SG17（安全研究组）负责制定的大数据安全相关标准包括编制中的《移动互联网服务中的大数据分析安全要求和框架》、《大数据即服务的安全指南》，《电子商务业务数据生命周期管理安全参考架构》等。

2.2.3 NIST

美国国家标准与技术研究院（NIST）于2012年6月启动了大数据相关基本概念、技术和标准需求的研究，2013年5月成立了NIST大数据公开工作组（NBG-PWG），2015年9月编写形成并发布了NIST SP 1500《NIST大数据互操作框架》系列标准（第一版），包括7个分册，即：NIST SP 1500-1《第1册 定义》、NIST SP 1500-2《第2册 大数据分类法》、NIST SP 1500-3《第3册 用例和一般要求》、NIST SP 1500-4《第4册 安全和隐私保护》、NIST SP 1500-5《第5册 架构调研白皮书》、NIST SP 1500-6《第6册 参考架构》和NIST SP 1500-7《第7册 标准路线图》。

其中，NIST SP 1500-4《NIST大数据互操作框架：第4册 安全与隐私保护》由NIST NBD-PWG的安全与隐私保护小组编写。

2.2.4 TC28

为推动和规范我国大数据产业的快速发展，培育大数据产业链，并与

国际标准接轨，全国信标委在2014年12月成立了大数据标准化工作组（以下简称“大数据工作组”，BDWG），工作组主要负责制定和完善我国大数据领域标准体系，组织开展大数据相关技术和标准的研究，推动国际标准化活动，对口ISO/IEC JTC1 WG9大数据工作组。目前，工作组正在制定的国家标准有12项，其中《信息技术 大数据 术语》等6项国家标准进入报批阶段，《信息技术 数据交易服务平台 交易数据描述》等3项标准进入征求意见阶段、1项标准完成草案，2项标准完成草案框架。

2.2.5 TC260

为了加快推动我国大数据安全标准化工作，全国信安标委在2016年4月成立大数据安全标准特别工作组（以下简称“特别工作组”，SWG-BDS），主要负责制定和完善我国大数据安全领域标准体系，组织开展大数据安全相关技术和标准研究。目前，特别工作组正在制定《信息安全技术 个人信息安全规范》、《信息安全技术 大数据服务安全能力要求》、《信息安全技术 大数据安全管理指南》等国家标准。其中，《信息安全技术 个人信息安全规范》和《信息安全技术 大数据服务安全能力要求》已经推进到征求意见稿阶段。同时，特别工作组组织开展了针对大数据安全能力成熟度模型、大数据交易安全要求、数据出境安全评估等国家标准的研究工作。

2.3 大数据安全相关标准现状

数据安全以数据为中心，重点考虑数据生命周期各阶段中的数据安全问题。大数据应用中包含海量数据，存在对海量数据的安全管理，因此，在分析大数据安全相关标准时，需要对传统数据采集、组织、存储、处理等安全相关标准进行适用性分析。此外，在大数据场景下，个人信息安全问题备受关注。由于大数据场景下的多源数据关联分析可能导致传统的个

人信息保护技术失效，因此，大数据场景下更需要考虑个人信息安全问题，必须对现有个人信息保护技术和标准进行适用性分析。最后，大数据应用作为一个特殊的信息系统，除存在与传统信息安全一样的保密性、完整性和可用性要求外，还需要从管理角度研究大数据场景下信息系统的安全，因此，传统信息系统的大部分信息安全管理体系和管理要求类标准仍然是适用的。下面对和大数据安全相关的传统数据安全标准、个人信息保护标准和专门为大数据应用制定的大数据安全相关标准进行梳理分析。

2.3.1 传统数据安全标准规范

（一）支付卡行业数据安全标准介绍

支付卡行业数据安全标准（PCI-DSS）是PCI安全标准委员会制定的数据安全标准。PCI-DSS标准目标在于严格控制对支付卡持卡人数据的处理、存储和传输，以保障银行卡用户在线交易的安全。PCI-DSS标准按每年交易量将商家分为四个等级，为不同等级商家提出不同强度的安全要求。PCI-DSS要求所有涉及信用卡支付的企业必须满足PCI-DSS标准。

PCI-DSS安全标准部分主要内容包括6大类要求：

1）构建和维护一个安全的网络；

2）保护持卡人数据；

3）维护一个脆弱性管理流程；

4）实施强制访问控制措施；

5）定期监控和测试网络；

6）维护一个信息安全策略。

对于每一类要求，PCI-DSS对其进行了详细规定，达到可操作的要求，比如，对于第1类安全要求（构建和维护一个安全的网络），其规定：

1）安全维护一个防火墙配置来保护持卡人数据；

2）不要在系统密码和其它安全参数方面使用默认值。

PCI-DSS也处于不断发展之中，比如，针对云计算新型环境，PCI-DSS制定了专门的补充标准《信息补充：PCI-DSS云计算指南》。

（二）NCHHSTP数据安全和私密性指南

美国艾滋病、肝炎、性传播疾病与结核病预防中心（NCHHSTP）发布了旨在实现HIV、病毒性肝炎、性传播疾病，和肺结核监护数据共享的数据安全和私密性指南。该指南详细分析了共享数据、维护安全和私密性的好处、风险和代价。给出了实现数据收集、存储、共享和使用过程安全和私密性的10大指导原则，并制定了实现数据收集、存储、共享和使用过程中安全和私密性的安全指南。

NCHHSTP数据安全10大原则为：

1）公共健康数据的获取、使用、披露和存储必须为合法公共健康目的服务；

2）应该只收集最小数量的个人识别数据以执行必要的公共安全活动；

3）必须拥有保护个人识别数据隐私和安全的强安全策略；

4）数据的收集和适用策略必须反映出对个人和社区组织的尊重，且要减轻他们不必要的负担；

5）必须有确保所采集和使用数据质量的策略和流程；

6）有责任及时使用和分发摘要数据给相关干系人；

7）数据共享应该只限于合法的公共健康目的，且必须及时为数据共享建立数据使用协议；

8）公共健康数据应该在一个安全环境中保存，以及通过安全方法传输；

9）最小化被授权访问可识别个人信息的人员和实体的数量；

10）职员应该主动负责对公共健康数据的管理。

NCHHSTP数据安全和隐私性指南包括五个方向：策略和责任、数据的收集和使用、数据的共享和发布、物理安全和电子数据安全。

2.3.2 个人信息安全标准规范

（一）ISO/IEC 29100:2011《信息技术　安全技术　隐私保护框架》

该标准为信息与通信技术（ICT）系统内可识别个人信息（PII）的保护提供了一个高层次隐私保护框架。该隐私保护框架规范了通用的隐私保护术语；定义了处理PII中的参与者及其角色；描述了隐私保护的考虑事项；为实现由许多国际组织开发的11个隐私保护原则提供指导。11个隐私保护原则包括同意和选择、意图合法性和规约、收集限制、数据最小化、使用/保留/披露限制、准确和质量、开放/透明/告知、个体参与和访问、可核查性、信息安全、隐私保护合规。该标准适用于涉及规范、获取、构建、设计、开发、测试、维护、管理和运行需要隐私保护控制措施来处理PII的ICT系统或服务的任何自然人和组织。

（二）ISO/IEC 29101:2013《信息技术　安全技术　隐私保护体系结构框架》

该标准定义了一个隐私参考体系结构框架，该框架明确提出了处理PII的ICT系统的关心点，列出了实现这种系统的组件，并提供了将这些组件语境化的体系结构视图。该标准适用于涉及规划、获取、构建、设计、测试、维护、管理和运行处理PII的ICT系统的实体。

（三）ISO/IEC 29190:2015《信息技术　安全技术　隐私保护能力评估模型》

该标准为组织评估其管理隐私保护相关过程的能力提供高层指南，规范了确定隐私保护能力的评估过程和评估级别，为评估隐私保护能力的关键过程域及其实现，以及如何将隐私保护能力评估继承到组织运行中提供

了指南。

（四）ISO/IEC 27018:2014《信息技术　安全技术　可识别个人信息（PII）处理者在公有云中保护PII的实践指南》

该标准依据ISO/IEC 29100给出的隐私保护原则，为在公有云计算环境中保护可识别个人信息（PII），建立了普遍接受的控制目标、控制措施和测量实现指南。特别是，该标准考虑到在公有云提供者的信息安全风险环境下适用的PII保护法规要求，基于ISO/IEC 27002给出指南。该标准适用于作为PII处理者通过云计算提供信息处理服务的所有类型和规模的组织。

（五）ISO/IEC 29134《信息技术　安全技术　隐私影响评估指南》

该标准为隐私影响评估（PIA）过程以及PIA报告的结构和内容给出指南。该标准适用于所有类型和规模组织。

（六）ISO/IEC 29151《信息技术　安全技术　可识别个人信息（PII）保护实践指南》

该标准为满足通过可识别个人信息（PII）保护相关的风险和影响评估而识别的要求，建立了控制目标和控制措施，并提供了控制措施实现指南。该标准考虑到在组织信息安全风险环境下适用的PII处理要求，基于ISO/IEC 27002给出指南。该标准适用于作为PII控制者的所有类型和规模的组织。

（七）BS 10012:2009《数据保护　个人信息管理系统规范》

该标准由英国标准协会（BSI）于2009年6月发布，主要是针对个人信息保护所提出的“个人信息保护标准”，其中参考了经济开发合作组织（OECD）的个人隐私权保护的八大原则，用于支撑欧盟隐私保护条例和英国的数据保护法案，强调要建立一个管理体系，并且就个人信息跨境管理的情况给出了建议。

该标准规范了个人信息管理体系（PIMS）要求，提供了一个框架用

于维护和改进数据保护的合规性和最佳实践。该标准适用于任何规模和行业的组织，主要为在其内部启动、实施和维护PIMS的组织所用。该标准旨在提供个人信息管理的共同基础，以便增强个人信息管理的信心，并使得内部和外部评估者能够有效地评估数据保护的合规性和最佳实践。

（八）《信息安全技术 个人信息安全规范》（国家标准，在研）

该标准提出了通过计算机系统处理个人信息时，应当遵循的原则和采取的安全控制措施。该标准要求个人信息控制者在使用计算机系统对个人信息进行处理时，应遵循以下基本原则：目的明确原则、同意和选择原则、最少够用原则、开放透明原则、质量保证原则、确保安全原则、个体参与原则、问责原则、披露限制原则。该标准用于指导组织内部建立个人信息保护策略，并用于指导产品、服务、内部信息系统的设计、开发和实现。

2.3.3 大数据安全标准规范

（一）ISO/IEC 20547–4《信息技术 大数据参考架构 第4部分：安全与隐私保护》（国际标准，在研）

该标准分析了大数据面临的安全与隐私保护问题和相关风险，在ISO/IEC 20547-3《信息技术 大数据参考架构 第3部分：参考架构》给出的大数据参考架构（BDRA）基础上，提出了大数据安全与隐私保护参考架构（BDRA-S&P）。BDRA-S&P包括用户视角的大数据安全与隐私保护角色和活动，以及功能视角的支持大数据安全与隐私保护活动的功能组件。该标准还汇集了信息安全领域中已有的安全控制措施和隐私保护控制措施，作为大数据安全与隐私保护功能组件的选项。

（二）NIST 1500–4《NIST大数据互操作框架：第4册 安全与隐私》（美国标准）

该标准聚焦于提出、分析和解决大数据特有的安全与隐私保护问题。

在理解和执行安全与隐私保护要求上，大数据触发了需求模式的根本转变，从而满足大数据的体量大、种类多、速度快和易变化的特点。基础架构的安全解决方案目标也发生了变化，例如，分布式计算系统和非关系型数据存储的安全。大数据场景下新的安全问题需要解决，其中包括平衡隐私与实用性，对加密数据开展分析和治理，以及核查认证用户和匿名用户。该标准分析了特定应用场景（包括医疗、政府、零售、航空等）下的大数据安全与隐私保护问题，提出了大数据安全与隐私保护的主要概念和角色，开发了一个大数据安全与隐私保护参考架构来补充NIST大数据参考架构（NBDRA），并对行业应用案例和NBDRA之间的映射进行了相关探索。

（三）《大数据服务安全能力要求》（国家标准，在研）

该标准定义了大数据服务业务模式、大数据服务角色、大数据服务安全能力框架和大数据服务的数据安全目标和系统安全目标，规范了大数据服务提供者的大数据服务基本安全能力、数据服务安全能力和系统服务安全能力要求，为大数据服务提供者的组织能力建设、数据业务服务安全管理、大数据平台安全建设和大数据安全运营规范安全能力要求。该标准一方面可以为大数据服务提供者提升大数据服务安全能力提供指导，另一方面则为第三方机构对大数据服务安全测评提供依据。

该标准将大数据服务安全能力分为一般要求和增强要求。大数据服务提供者应依据大数据框架服务模式和大数据应用模式，根据大数据系统所存储和分析数据的敏感度和业务重要性的不同，提供相应级别的大数据服务安全能力。

（四）《大数据安全管理指南》（国家标准，在研）

该标准分析了数据生命周期各阶段中的主要安全风险，尤其是在数据转移的环节，对角色提出安全管理要求。该标准指导大数据生态环境中各角色安全地管理和处理大数据，形成一个安全大数据环境，确定各角色的

责任和行为规范，为各角色安全地处理大数据提出管理和技术要求。该标准规范大数据处理中的各个关键环节，为大数据应用和发展提供安全的规范原则，解决数据开放、共享中的基本原则。

第三章 大数据安全标准体系

3.1 大数据安全挑战

大数据安全风险伴随大数据应用而生。随着互联网、大数据应用的爆发，数据丢失和个人信息泄漏事件频发，地下数据交易黑灰产造成数据滥用和网络诈骗，并引发恶性社会事件，甚至危害国家安全。如2015年5月，美国国税局宣布其系统遭受攻击，约71万人的纳税记录被泄露，同时约39万个纳税人账户被冒名访问；2016年8月，犯罪团伙利用非法获取得到的数万条高考考生信息实施诈骗，山东女孩徐玉玉因学费被骗出现心脏骤停，最终抢救无效死亡；2016年12月，雅虎公司宣布其超过10亿的用户账号已被黑客窃取，相关信息包括姓名、邮箱口令、生日、邮箱密保问题及答案等内容；2016年至今，全球范围内数以万计的MongoDB系统遭到攻击，大量系统被黑客索取赎金。

通过对当前典型大数据应用场景以及大数据产业发展现状进行调研分析，本白皮书从技术平台和数据应用两个角度来讨论当前大数据发展面临的安全挑战。

3.1.1 技术平台角度

伴随着大数据的飞速发展，各种大数据技术层出不穷，新的技术架构、支撑平台和大数据软件不断涌现，使得大数据也面临着新的安全挑战。

（一）传统安全措施难以适配

海量、多源、异构、动态性等大数据特征导致其与传统封闭环境下的数据应用安全环境有所区别。大数据应用一般采用底层复杂、开放的分布式计算和存储架构为其提供海量数据分布式存储和高效计算服务，这些新

的技术和架构使得大数据应用的网络边界变得模糊，传统基于边界的安全保护措施不再有效。同时，新形势下的高级持续性威胁（APT）、分布式拒绝服务攻击（DDoS）、基于机器学习的数据挖掘和隐私发现等新型攻击手段出现，也使得传统的防御、检测等安全控制措施暴露出严重不足。

（二）平台安全机制亟待改进

现有大数据应用中多采用通用的大数据管理平台和技术，如基于Hadoop生态架构的HBase/Hive、Cassandra/Spark、MongoDB等。这些平台和技术在设计之初，大部分考虑是在可信的内部网络使用，对大数据应用用户的身份鉴别、授权访问、密钥服务以及安全审计等方面考虑较少。即使有些软件做了改进，如增加了Kerberos身份鉴别机制，但整体安全保障能力仍然比较薄弱。同时，大数据应用中多采用第三方开源组件，对这些组件缺乏严格的测试管理和安全认证，使得大数据应用对软件漏洞和恶意后门的防范能力不足。

（三）应用访问控制愈加复杂

由于大数据数据类型复杂、应用范围广泛，它通常要为来自不同组织或部门、不同身份与目的的用户提供服务。一般地，访问控制是实现数据受控访问的有效手段。但是，由于大数据应用场景中存在大量未知的用户和数据，预先设置角色及权限十分困难。即使可以事先对用户权限分类，但由于用户角色众多，难以精细化和细粒度地控制每个角色的实际权限，从而导致无法准确为每个用户指定其可以访问的数据范围。

3.1.2 数据应用角度

大数据的一个显著特点是其数据体量巨大，而其中又蕴含着巨大的价值。数据安全保障是大数据应用和发展中必须面临的重大挑战。

（一）数据安全保护难度加大

大数据拥有巨大的数据，使得其更容易成为网络攻击的显著目标。在

开放的网络化社会，蕴含着海量数据和潜在价值的大数据更受黑客青睐，近年来也频繁爆发信息系统邮箱账号、社保信息、银行卡号等数据大量被窃的安全事件。分布式的系统部署、开放的网络环境、复杂的数据应用和众多的用户访问，都使得大数据在保密性、完整性、可用性等方面面临更大的挑战。

（二）个人信息泄漏风险加剧

由于大数据系统中普遍存在大量的个人信息，在发生数据滥用、内部偷窃、网络攻击等安全事件时，个人信息泄漏产生的后果将远比一般信息系统严重。另一方面，大数据的优势本来在于从大量数据的分析和利用中产生价值，但在对大数据中多源数据进行综合分析时，分析人员更容易通过关联关系挖掘出更多的个人信息，从而进一步加剧了个人信息泄漏的风险。

（三）数据真实性保障更加困难

大数据系统中的数据来源广泛，可能来源于各种传感器、主动上传者以及公开网站。除了可信的数据来源外，同时存在大量不可信的数据来源。甚至有些攻击者会故意伪造数据，企图诱导数据分析结果。因此，对数据的真实性确认、来源验证等需求非常重要。然而，由于采集终端性能限制、技术不足、信息量有限、来源种类繁杂等原因，对所有数据进行真实性验证存在很大的困难。

（四）数据所有者权益难以保障

大数据应用过程中，数据会被多种角色用户所接触，会从一个控制者流向另外一个控制者，甚至会在某些应用阶段挖掘产生新的数据。因此，在大数据的共享交换、交易流通过程中，会出现数据拥有者与管理者不同、数据所有权和使用权分离的情况，即数据会脱离数据所有者的控制而存在，从而会带来数据滥用、权属不明确、安全监管责任不清晰等安全风险，将严重损害数据所有者的权益。

3.2 大数据安全标准化需求

大数据安全标准是应对大数据安全需求的重要抓手。基于对上面大数据安全风险和挑战的综合分析，以及对当前大数据技术和应用发展现状，以及当前我国对大数据安全合规方面的要求，提出五个方面的大数据安全标准化需求。

（一）规范大数据安全相关术语和框架

当前，大数据技术和应用在快速变化之中，人们对一些大数据概念和术语的认知水平不同，包括大数据定义、大数据安全角色、大数据生命周期等，所有这些都将影响大数据行业的快速和健康发展；同时，当前缺乏一个通用的能够清晰描述大数据生态中各安全角色之间关系和以及各角色安全活动的安全参考框架，以指导后续大数据安全标准的制定。因此，应优先制定包括大数据安全概念和框架、角色和模型等基础标准，为其它标准的制定打好坚实基础。

（二）为大数据平台安全建设、安全运维提供标准支撑

大数据平台和应用是支撑数据收集、传输、存储、处理和共享等数据活动的分布式信息系统，它包括底层的基础平台和上层的大数据应用。大数据平台和应用的安全建设和安全运维对整个大数据系统的安全产生重要影响。但当前，我国缺乏针对大数据基础平台和上层大数据应用的安全规范和指南，覆盖管理、工程、技术、平台系统和应用服务等各个方面，以指导大数据系统所有者、建设者、运营者对大数据平台和应用的安全建设、安全运维和安全风险管理。

（三）为数据生命周期管理各个环节提供安全管理标准

数据是大数据系统中的重要资源，其安全性至关重要。当前我国缺乏针对大数据环境下的数据管理安全规范，需要制定规范大数据系统中的数据安全管理活动、流程和方法的安全标准，以指导数据控制者的数据生命

周期管理活动，包括数据收集、传输、存储、共享、处理、共享等安全活动，减少来自组织内部和外部的各种大数据安全风险。

（四）为大数据服务安全管理提供安全标准支撑

大数据服务可以为大数据生态中的数据提供者和数据消费者提供数据分析处理、数据交易等服务。在提供大数据服务的过程中，大数据服务组织的安全能力至关重要，它直接影响到数据的安全性。当前，我国缺乏指导组织建立大数据服务安全能力的规范，以及对大数据服务组织的安全能力成熟度进行评级的标准规范，需要制定相关的标准，规范大数据服务组织的基础安全能力、数据安全管理能力和系统安全建设、安全运维能力，以及对组织安全能力成熟度进行有效评价，并指导其安全能力提升的标准。目前，我国大数据交易服务安全面临没有标准规范的局面，亟需建立大数据交易服务相关安全标准规范，支撑《网络安全法》在大数据交易领域的落地实施，为提升对大数据交易服务安全的管控能力，促进大数据交易服务产业安全健康发展提供标准依据。

（五）为行业大数据应用的安全和健康发展提供标准支撑

不同行业和领域的大数据应用具有不同特点，所涉及的数据敏感度因政策环境、行业环境不同存在差异，需要制定相应的行业大数据安全标准规范我国的行业大数据安全：一是在构建大数据安全标准体系时，统筹考虑数据在行业之间或组织之间的交换与共享问题，支撑行业大数据应用的快速发展；二是在标准制定层面，需要对电子政务、电子商务、电信、健康医疗等重点行业大数据应用适时出台相应的大数据安全指南类标准，指导各行业的大数据安全建设和运营。

3.3 大数据安全标准体系框架

基于国内外大数据安全实践及标准化现状，参考大数据安全标准化需求，结合未来大数据安全发展趋势，构建了如图3-1所示的大数据安全标

准体系框架。该标准体系框架由五个类别的标准组成，分别为：基础类标准、平台和技术类标准、数据安全类标准、服务安全类标准和行业应用类标准。

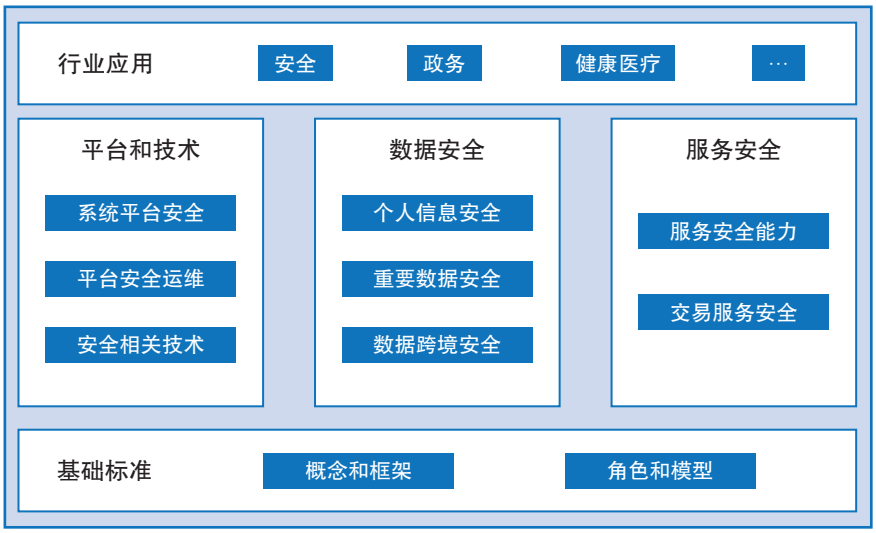


图3–1 大数据安全标准体系框架

（一）基础类标准

为整个大数据安全标准体系提供包括概述、术语、参考架构等基础标准，明确大数据生态中各类安全角色及相关的安全活动或功能定义，为其它类别标准的制定奠定基础。

（二）平台和技术类标准

该类标准主要针对大数据服务所依托的大数据基础平台、业务应用平台及其安全防护技术、平台安全运行维护及平台管理方面的规范，包括系统平台安全、平台安全运维和安全相关技术三个部分。系统平台安全主要涉及基础设施、网络系统、数据采集、数据存储、数据处理等多层次的安全技术防护。平台安全运维主要涉及大数据系统运行维护过程中的风险管

理、系统测评等技术和管理类标准。安全相关技术主要涉及分布式安全计算、安全存储、数据溯源、密钥服务、细粒度审计等安全防护技术。

（三）数据安全类标准

该类标准主要包括个人信息、重要数据、数据跨境安全等安全管理与技术标准，覆盖数据生命周期的数据安全，包括分类分级、去标识化、数据跨境、风险评估等内容。

（四）服务安全类标准

该类标准主要是针对开展大数据服务过程中的活动、角色与职责、系统和应用服务等要素提出相应的服务安全类标准，包括安全要求、实施指南及评估方法；针对数据交易、开放共享等应用场景，提出交易服务安全类标准，包括大数据交易服务安全要求、实施指南及评估方法。

（五）行业应用类标准

该类标准主要是针对重要行业和领域大数据应用，对涉及国家安全、国计民生、公共利益的关键信息基础设施的安全防护，形成面向重要行业和领域的大数据安全指南，指导相关的大数据安全规划、建设和运营工作。

3.4 大数据安全标准规划

根据大数据安全标准体系框架，通过对大数据基础标准、平台和技术、数据安全、服务安全、行业应用五个类别的标准需求梳理，明确了大数据安全标准化需求，通过对已发布及在研大数据安全相关标准的适用性分析和大数据安全标准缺口分析，编制了如图3-2所示的大数据安全标准规划，为我国近期的大数据安全标准的制修订提供指引。由于大数据技术和应用仍然处于快速演变之中，对于还未达到一定成熟度的可标准化的大数据安全主题，暂时不在本标准规划中列出，后续可以根据标准化需求不断对该标准规划进行补充。

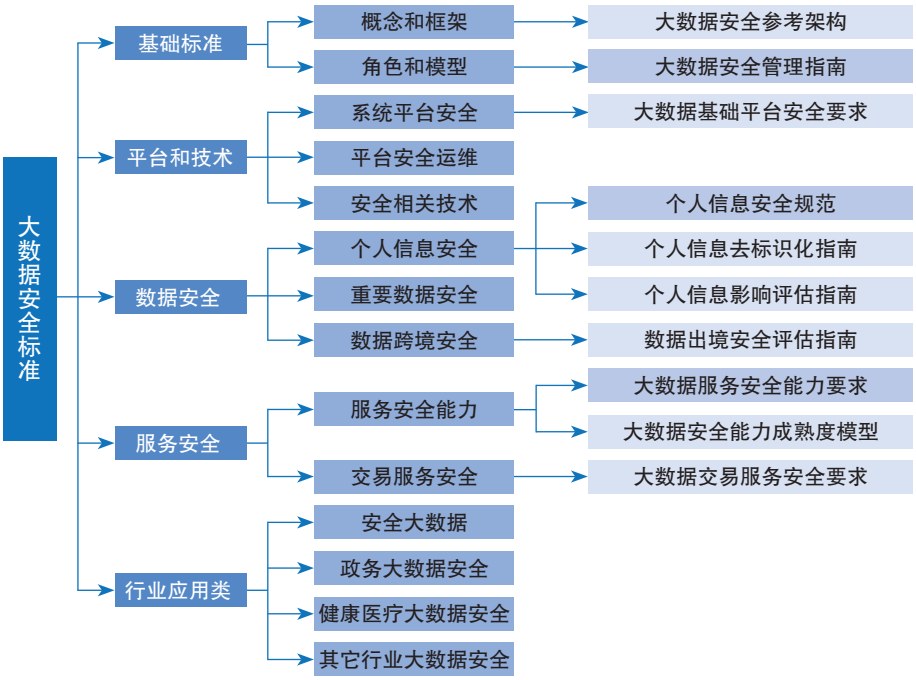


图3–2 大数据安全标准规划

关于标准规划中各大数据安全标准的简要内容说明详见表3-1。

表3–1 在研和建议研制的大数据安全相关标准说明

标准名称	状态	内容概述
大数据安全参考架构	建议研制	本标准给出了一个大数据安全参考模型，作为对大数据参考模型的重要补充，明确大数据平台和应用应提供的安全功能组件，以及安全组件之间安全接口，为其它大数据安全标准提供基础支撑。
大数据安全管理指南	在研	本标准明确了大数据生态中各安全角色及安全责任，建立大数据安全管理模型，围绕数据生命周期管理各阶段，提出安全控制措施指南。
大数据基础平台安全要求	建议研制	本标准的标准化对象为大数据框架提供者所构建的大数据基础平台，规范大数据基础平台的各项安全技术要求，如大数据平台的安全防御、检测方面的技术要求。

标准名称	状态	内容概述
个人信息安全规范	在研	本标准提出了通过计算机系统处理个人信息时，应当遵循的原则和采取的安全控制措施。本标准用于指导组织内部建立个人信息保护策略，并用于指导产品、服务、内部信息系统的设计、开发和实现。
个人信息去标识化指南	建议研制	本标准提出个人信息去标识化的具体指导，包括原则，方法和流程。用于指导个人信息控制者开展个人信息去标识化工作。本标准在平衡数据可用性和个人信息安全的前提下促进数据的开放、共享和交易。
个人信息影响评估指南	建议研制	本标准提出了个人信息影响评估的原则、方法、流程，用于指导组织评估个人信息处理活动对个人信息主体合法权益可能造成的影响，并指导组织采取必要的措施降低不利影响的风险。本标准旨在确保个人信息保护的基本原则能有效地在新技术、新商业模式中得到有效落实。
数据出境安全评估指南	建议研制	本标准规定了数据跨境流动安全评估指标及评估办法，使政府及企业自身可以对数据跨境流动安全进行全面评估。适用于各类组织开展的数据出境安全评估工作，也适用于网络安全相关主管部门、第三方评估机构等组织开展数据出境安全评估、监督管理等工作。
大数据服务安全能力要求	在研	本标准规范了大数据服务提供者的在提供服务时应该具备的基本安全能力、数据服务安全能力和系统服务安全能力。可为大数据服务提供者提升大数据服务安全能力提供指导，同时可为第三方机构对大数据服务安全能力测评提供依据。
大数据交易服务安全要求	建议研制	本标准旨在规范大数据交易服务平台的安全交易流程和安全要求，包括安全技术和安全管理方面的要求，在保障数据交换和共享过程中数据的安全的同时，保护数据挖掘利用过程中的数据安全和个人信息安全要求。旨在减少现有大数据交易中出现的各种安全问题，包括交易中的信息泄密、数据滥用和个人信息泄露等问题。
大数据安全能力成熟度模型	建议研制	本标准旨在帮助大数据组织建立一套评价数据安全管理、系统安全建设和运维能力的通用术语和成熟度评价模型，指导大数据组织建立数据安全能力评价模型及提升数据安全能力的方案，为第三方机构评价组织的数据安全成熟度水平提供了依据基准，以促进大数据行业的健康发展和公平竞争。

第四章 大数据安全标准化工作建议

4.1 加强大数据安全核心技术研究

发展大数据安全核心技术可以有效促进大数据产业自主发展，并可支撑自主大数据安全标准的制定和落地实施。当前的信息安全技术并不能满足大数据应用的安全需求，需要对现有信息安全技术进行完善，或研究新的大数据安全技术，以解决大数据应用过程中的各种特有安全风险。建议加强大数据安全技术研究，包括分布式环境下的数据完整性验证、数据标签、区块链、细粒度访问控制、密文透明运算、数据溯源、数据脱敏与安全审计等技术；同时，建议加强使用大数据安全技术来进行网络安全入侵检测、安全态势感知、网络攻击取证、威胁情报分析等安全应用研发，以利用大数据技术来抵御针对大数据的攻击威胁。

4.2 加快制定个人信息安全相关标准

近年，随着互联网经济、互联网社交等新业态的普及，越来越多的系统和平台收集个人信息，并对个人信息进行存储、处理，甚至交换。个人信息的非法收集、泄露、滥用等已成为社会性关注的焦点问题，个人权益严重受损情况屡见不鲜，甚至出现了很多与个人信息滥用有关的违法犯罪活动。为了支撑《网络安全法》的落地，亟需制定个人信息保护相关的安全标准，定义个人信息和个人敏感信息，规范个人信息收集、存储、处理、使用和披露等各个环节中数据操作的相关行为，并制定配套的风险评估和技术实现支撑标准，对个人信息处理生命周期中的风险进行评估并提供个人信息去标识化的技术、方法和流程。旨在遏制个人信息滥用乱象，最大程度地保障用户合法权益和社会公共利益。

4.3 加快制定数据共享相关安全标准

数据作为一种战略资源，在开放和共享过程中会产生更大价值。但数据共享过程中由于安全技术缺乏或安全管理能力不足暴露了很多安全问题，包括地下数据黑灰产，以及共享过程中侵犯用户个人信息安全、数据滥用等问题，严重阻碍了数据共享进程。建议建立与数据共享相关的数据安全管理办法，加快数据交易安全相关标准的制定工作，规范数据交易市场，从数据交易平台、交易主体、交易对象、交易过程等方面规范数据交易服务安全；加快大数据安全能力成熟度模型相关标准的制定工作，规范组织机构在大数据环境下开展数据业务时应具备的数据安全能力，用于评估组织机构的大数据安全保障能力，确保数据在流转过程中得到了充分保护，有效解决数据共享中的各种安全问题，保障数据交易各方的利益，促进大数据产业的健康和安全发展。

4.4 加快制定数据出境安全相关标准

为降低跨境数据流动潜在的安全风险，保障国家安全、公共利益和公民权益，《网络安全法》明确规定，关键信息基础设施运营者因业务需要，向境外提供在中国境内运营中收集和产生的个人信息和重要数据的，需按照国家网信部门会同国务院有关部门制定的办法进行安全评估。为尽快落实有关规定，亟需制定数据跨境的相关安全标准。明确研究数据出境安全评估的主要风险指标、数据属性特征指标，判断出境数据的重要性，设计出境活动评估指标，综合评判出境活动的风险性。为国家开展数据出境安全评估审查的工作机制和有关制度落地提供标准支撑，为企业开展数据跨境安全风险自评估提供规范指南。

4.5 加快大数据安全审查支撑性标准研制

为提高大数据产品和服务的安全可控水平，防范大数据应用中的各种数据安全和隐私安全风险，维护国家安全和公众利益，依据《网络安全法》和《网络安全产品和服务审查办法》，我国将对影响国家安全和公众利益的大数据系统和服务进行安全审查。但当前缺乏对大数据安全审查的支撑性相关标准，亟需制定相关国家标准，对大数据服务提供者的数据管理能力、服务安全能力、组织安全能力、安全合规程度等进行审查，为第三方审查机构的安全审查工作提供重要支撑。

4.6 大力推广大数据安全标准应用

加快《个人信息安全规范》、《大数据服务安全能力要求》等标准的验证、试点示范工作，提高大数据安全标准的有效性和可操作性，为标准推广实施做好技术储备，积累实施经验。选取若干试点企业，开展标准适用性和实施效果评价，跟踪标准使用情况，发现问题，总结经验，完善标准，并更好地指导下一步大数据安全标准的立项、制定和推广工作。加大大数据安全标准宣传培训力度和推广实施，围绕重点领域组织编写大数据安全标准实施和应用指南，联合网络安全主管部门、高校、人才和创新基地等，开展大数据安全标准的宣传培训和推广实施。

4.7 加强大数据安全标准化人才的培养

人才是落实大数据安全保障的重要力量。建议建立健全多层次、多类型的大数据安全人才培养体系，鼓励高校、企业、测评机构等单位合作，在网络空间安全学科下设立大数据安全课程，重点培养大数据安全技术人才，以及大数据安全标准制定、宣贯、检测、评估类专业人才。鼓励高校和企业合作，实行大数据安全综合型人才联合培养模式。依托社会化教育资源，开

展大数据安全知识普及和教育培训，提高社会整体大数据安全认知水平。

4.8 深度参与大数据安全国际标准制定

紧密跟踪研究国际国外大数据安全标准化发展趋势和工作动态，编制大数据安全标准化白皮书等报告，做好标准化支撑工作，加强大数据安全国际标准提案研究，加大对我国单位和专家在大数据安全国际标准项目中担任编辑并主导编制的工作支持力度，深度参与大数据安全国际标准编制工作。充分发挥现有国际标准化交流与合作机制的优势，举办大数据安全标准化国际交流合作活动。推动大数据安全领域国际标准提案，将国内成熟的大数据安全标准转化为国际标准，贡献中国智慧，提升我国在大数据安全国际标准制定方面的国际话语权和影响力。

缩略语

APEC	亚太经济合作组织（Asia-Pacific Economic Cooperation）
BDWG	大数据标准工作组（Big Data Working Group）
BDRA	大数据参考架构（Big Data Reference Architecture）
BDRA-S&P	大数据参考架构（Big Data Reference Architecture — Security and Privacy）
BSI	英国标准协会（British Standard Institute）
DPI	深度包检测（Deep Packet Inspection）
ePHI	电子受保健康信息（electronic Protected Health Information）
ETI	欧洲透明度倡议（European Transparency Initiative）
FERPA	[美国]家庭教育权和隐私权法案（Family Educational Rights and Privacy Act of 1974）
GDPR	[欧盟]通用数据保护条例（General Data Protection Regulation）
HIPAA	[美国]健康保险携带和责任法案（Health Insurance Portability and Accountability Act）
ICT	信息与通信技术（Information and Communications Technology）
IEC	国际电工委员会（International Electrotechnical Commission）
ISO	国际标准化组织（International Organization for Standardization）
ITU-T	国际电信联盟电信标准化组（International Telecommunication Union - Telecommunication Standardization Sector）

JTC1	ISO/IEC联合技术委员会1：信息技术（ISO/IEC Joint Technical Committee 1: Information technology）
NBDRA	NIST大数据参考架构（NIST Big Data Reference Architecture）
PCI DSS	支付卡行业数据安全标准（Payment Card Industry Data Security Standard）
PIA	隐私影响评估（Privacy Impact Assessment）
PII	可识别个人信息（Personally Identifiable Information）
PIPC	[日本]个人信息保护委员会（Personal Information Protection Committee）
PHI	受保护健康信息（Protected Health Information）
NBD-PWG	NIST大数据公开工作组（NIST Big Data Public Working Group）
NCHHSTP	[美国]艾滋病、肝炎、性传播疾病与结核病预防中心（National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention）
NIST	[美国]国家标准与技术研究院（National Institute of Standards and Technology）
SWG-BDS	大数据安全标准特别工作组（Special Working Group - Big Data Security）

附录A

典型行业大数据应用和安全风险

1. 安全大数据

网际空间安全面临的威胁越来越多样化。移动网络、云和虚拟化、物联网、工控系统等技术领域的快速发展，使得保护对象和攻击路径都变得更加复杂。而攻击来源也从早期的个人黑客变为犯罪团伙、政治势力、网络部队等更严密的组织。甚至大数据技术本身也被攻击者所利用。能够应对核威慑的，只有核威慑本身；能够应对大数据攻击技术的，也只有大数据安全技术。目前安全行业的大数据应用场景主要包括等几类：

（一）网络安全态势感知

近年来，网络安全事件层出不穷，传统安全防御措施很难及时、有效的发现安全威胁。这就需要依靠互联网的海量安全数据，解决网络安全监控的问题，通过大数据技术对这些安全要素信息进行分析，全面、精准的掌握网络安全状态，并以可视化的方式，向网络安全监管单位提供所属管辖范围内的实时感知，同时针对安全隐患提供通报等手段帮助监管单位完成安全监控的闭环，从而改变当前“黑客主动攻击、企业被动防御”的局面。

态势感知技术这一概念源于美国空军的研究，此后在核反应控制、空中交通监管及医疗应急调度等领域被广泛应用。在安全领域，该技术是指广泛采集和收集广域网中的安全状态和事件信息，并加以处理、分析和展现，从而明确当前网络的总体安全状况，为大范围的预警和响应提供决策支持的技术。态势感知技术主要是应对大范围广谱威胁，相关的技术包

括海量异构数据分析、深度学习、网络综合度量指标、网络测绘、威胁情报、知识图谱、安全可视化等。

（二）高级持续威胁检测

高级持续性威胁具有精心伪装、定点攻击、长期潜伏、持续渗透等特点，已经成为网络犯罪和间谍活动的首选攻击方式。过去针对特定网络APT定向攻击的发现有两个难点：一是未知威胁分析过程缺少对历史数据的支持，难以进行回溯关联，遗漏了很多关键信息；二是缺少外部情报的来源，只依赖于自有的黑域名/黑IP库，检测的精度和效率都难以满足需求。

采用大数据技术，从两方面搜集数据：一是来自于互联网威胁情报云平台的威胁情报数据，二是来自于本地运营商互联网出口监控到的网络流量数据。基于上述的海量安全数据，可以通过人工智能结合大数据知识以及攻击者的多个维度特征还原出攻击者的全貌，包括程序形态，不同编码风格和不同攻击原理的同源木马程序，恶意服务器等，通过全貌特征跟踪攻击者，持续地发现未知威胁。通过对云端大数据中提取的恶意域名、IP、主防库、样本库等信息进行关联分析，发现传统规则检测手段无法发现的未知威胁，实现攻击早期的快速发现。根据样本外连行为识别免杀木马，实现早期的快速发现。对未知威胁的网络行为，攻击源头进行精准定位，对远控木马等行为进行威胁的识别，最终达到对入侵途径及攻击者背景的研判与溯源。

（三）伪基站发现与追踪

伪基站是一种小型或微型的信号收发装置，和运营商的真实基站类似，能够获取周围的手机与基站的设备信息，通过模拟真实基站通信机制，迫使周围的手机连接到该仿冒的基站上，向普通用户发送垃圾短信，甚至冒用号码，群发诈骗信息。

采用大数据技术，则可以极大提高发现伪基站的能力和效率，并可以及时阻断诈骗短信中的钓鱼链接，打破诈骗链条。具体包括以下步骤：第

一，通过手机用户举报垃圾短信，或者通过手机专业软件主动拦截并上报垃圾短信，大量收集伪基站短信中包含的时间、地点、内容、仿冒的基站号等各种信息；第二，在大数据处理平台中，运用自然语言处理与机器学习的方法，去掉大量的噪声点，从海量的垃圾短信中以较高的精度识别出伪基站短信；第三，将伪基站短信与经纬度信息结合，就可以发现并定位伪基站；结合伪基站的历史数据，可以进一步找到伪基站的活动规律，并以此对其运动轨迹进行预判；第四，与地理信息系统联动，展现伪基站位置、伪基站的行为、历史运行路径、数量分布等等信息，从而帮助执法部门的抓捕活动。

（四）反钓鱼攻击

钓鱼攻击是一种利用社会工程学手段，伪装在线金融或交易平台的网站，针对客户个人身份数据和金融账号进行盗窃的犯罪行为。近些年来，钓鱼攻击相关的网银欺诈案件使得用户蒙受巨大的经济损失，也严重影响了银行业金融机构的声誉。据中国反钓鱼网站联盟发布的统计数据，在2015年9月，处理的钓鱼网站达1531个，涉及淘宝网、工商银行、平安银行、 建设银行四家单位的钓鱼网站总量占全部举报量的 98.56%。据《2014年中国网络购物安全报告》报告：在2014年，包括钓鱼攻击、恶意代码在内的安全威胁，给国内网购用户带来了超过300亿的损失。

发现钓鱼网站，需要利用搜索引擎扫描相关互联网址，并通过大数据建模过滤掉可信页面与重复页面，筛选出有嫌疑的钓鱼网址页面，将这些页面输入到分析引擎中；用户也可以进行举报，将钓鱼网址上报到分析引擎的数据库中。分析引擎通过规则模型综合研判、机器学习等方式检测出钓鱼网址和页面。将发现的钓鱼网站和网页汇集成为网址信誉库，金融机构可以把具有欺诈性的URL信息提到这个信誉库中，其中的信息就是是否拦截网页访问的依据。各终端访问钓鱼网站时，通过与云关联的终端软件，提示并阻止用户的访问行为。

2. 电子政务大数据

各政务部门在开展业务的过程中积累了大量数据，通过对政务大数据进行开发、利用、挖掘，能够为政府提供智能办公、智能监管、智能服务、智能决策等大数据服务，提高政府办公、监管、服务、决策的智能化水平，推动社会治理体系和治理能力现代化。通过将各级政务部门数据资源的汇集，实现政务数据的融合互通，并对大量的多源异构数据融合进行大数据综合分析、挖掘，帮助政府将现有的数据资源转化创造价值，能够实现政务大数据在城市规划、建设、管理等方面的综合应用。

（一）资源共享与政务协同

利用大数据技术整合各类政务信息资源、共享政务数据，提高行政效率、提升公共服务水平。构建大数据平台对接金税、金质、金盾、金卡、公共卫生系统等相关部门信息系统，实现跨部门、跨行业、跨平台的数据交换，同时可以基于国家、省级等数据中心建设，实现政务数据资源汇聚、共享，消除“数据割据”、“信息孤岛”，促使政务活动更加开放、透明，有效改善政府内部、政府部门之间，以及政府与其他管理主体之间的协同管理，提高政府与企业、政府与社会公众之间的政务协同和公共服务能力等。

（二）决策管理

利用大数据技术为政务决策提供支持，开展在政府应急决策管理、公共突发卫生事件决策、公共交通指挥决策、综合社会管理决策、环境污染保护决策等方面预测预警作用，能提升行政效率、优化社会公共服务。推动电子政务与城市管理、经济发展、民众保障、公共服务等多方面的深度融合，提供涵盖水电煤气、城建规划、交通物流、金融投资、教育就业、社保民生、旅游文化等行业领域的服务。

（三）精准扶贫

利用大数据精准确定对象，通过运用大数据技术，可以为政府建立起电子政务大数据信息的共享和交换平台，将上报的贫困户信息与民政、人社、银行、工商、车管、编办等部门的数据进行比对，为扶贫有关决策提供支持；同时，通过大数据对比和分析，对民政、人社、银行等信息进行综合分析，可以找出致贫原因，精准确定出谁是真正的贫困户，为精准帮扶提供依据。

（四）数据管税

利用大数据技术比对国土、人社、房产、车管等部门数据，通过互联网采集、大数据分析企业用水、用工、用电等数据，能够杜绝逃税漏税现象发生，实现税务数据的最大整合、涉税数据的互联互通、风险管理的全程监控。通过大数据共享平台，建立完善政府各部门数据共享机制，为各部门提供数据服务，用“大数据”发现问题、分析问题、解决问题，让“数据说话”、“综合治理”，提升政府决策分析水平，提高政府综合服务能力。

（五）智慧党建

目前，党员流动问题、底数不清、管理乏力、服务滞后等问题在农村特别严重，基本上年轻的党员都在外地，这就对党员管理就带来挑战，传统的党员管理手段，已经不能满足现在需求。通过构建智慧党建平台，采集和汇聚党员数据和信息，可以清晰的看到党组织分布情况、党支部性质、结构，以及党员的基本信息，利用大数据技术，可以分析和展现流动党员的具体位置、联络员、帮扶人员等情况，实现对党员的精准管理，构建全方位、立体式的党建工作体系。

政务大数据覆盖行业范围广泛、数据结构多样、关联关系复杂，而且涉及大量个人隐私数据、国家敏感数据等重要数据，因此在开展政务大数据应用的同时，数据和平台安全尤为重要。电子政务大数据面临的安全风险和挑战主要包括：

1）平台安全

大数据平台是政府使用数据资源的基础平台，平台安全是保障政府安全可靠利用数据资源的基础。大数据平台除了面临传统的恶意代码、攻击软件套件、物理损坏与丢失等安全威胁外，由于自身架构要根据政府业务需求和安全要求变化不断改进，因而产生传统的身份认证、数据加密手段适用性问题。

2）服务安全

构建基于互联网的一体化公共服务平台，面向公众提供基于大数据的便民服务，是落实国家推进国家治理体系和治理能力现代化、建设服务型政府要求的重要任务。基于互联网建设的政务在线服务窗口，是政务大数据为社会公众服务的重要组成部分，便捷的互联网应用环境下，在提质增优公共服务的同时也为便民服务带来严峻的安全挑战，需要应对基于 Web 的攻击、Web 应用程序攻击/注入攻击、拒绝服务攻击、网络钓鱼、用户身份盗窃等威胁，抵御信息泄露、网络瘫痪、服务中断等安全风险。

3）数据安全

各部门在开展业务和对政务大数据进行开发利用的同时，数据自身安全非常重要，涉及数据生命周期各阶段相关的数据采集、数据传输、数据存储、数据处理、数据交换、数据销毁等活动。政府部门数据公开、行业间以及行业内部数据平台化共享时的数据安全，是迫切需要解决的问题，是大数据资源实现开放共享、相关“数据掘金”应用得以发展的关键。

4）数据确权问题

政务数据的所有权、使用权、管理权涉及多个部门，特别是政府授权社会资本方搭建的公共服务系统所产生的数据，涉及个人隐私、国家经济命脉，在进行大数据分析中，必须做到权责分明，厘清数据权属关系，防止数据流通过程中的非法使用，保障数据安全流通。但是，目前数据权属仍缺乏法律支撑，数据使用尤其跨境流动所产生的安全风险日益凸显。

5）APT 攻击防御

APT是黑客针对客户所发动的网络攻击和侵袭行为，是一种蓄谋已久的“恶意网络间谍威胁”。这种行为往往经过长期的经营与策划，并具备高度的隐蔽性。APT攻击以窃取核心资料为目的，对政府部门大数据应用产生重大安全威胁，因此必须在政务大数据中高度防范此类攻击。

3. 健康医疗大数据

作为典型的实践科学，医学中有很多知识来源于经验积累。而目前经验积累的最直接、客观的体现就是“数据”。因此，利用健康医疗过程中产生的海量数据，开发其潜在价值，使其助力健康医疗事业的发展，成为医疗行业、技术研发领域等相关有识之士共同努力的目标。

（一）医学研究效率提升

中国作为世界上病例数最多的国家，为医生科研提供了天然的优势条件，但传统的数据存储方式，使得病例数据的收集成为科研最大的障碍。基于整合后的医疗大数据，利用大数据分析及搜索引擎技术可以有效延展研究范围和深度，帮助医生开展过去难以开展的研究领域：临床科研应用、跨医疗机构间的多中心联合研究、专科专病数据挖掘积累及研究，提升科研质量和效率。

（二）医院管理提效提能

解决医院管理数据质量差、口径杂、系统多、方法少等情况导致的医院管理者无法详细了解医院运营情况的问题，能为医疗机构职能部门提供日常指标的快捷多维度分析功能，提升管理人员的宏观洞察力。

（三）健康医疗生态良性发展

基于清洗挖掘后的标准化医疗大数据，还可以提供更多应用服务，如临床辅助决策服务（CDSS）、精准保险服务、药品研发等，大数据可助其节省研发成本、缩短研发周期、节省营销成本、提升营销精准率。

健康医疗大数据在促进业务发展的同时，面临的安全挑战主要表现在：

1）数据权属不清。健康医疗大数据起源于个人患者本身，那么数据权属到底是属于个人、还是产生数据的医疗机构一直没有定论；另外，第三方机构在原始数据基础上挖掘延伸出的新数据，其归属权也没有明确规定。

2）应用复杂性高。目前各地区和机构在进行健康医疗领域信息化建设时大都根据自身需求建立独立的信息系统，这些信息系统架构各异、数据格式不同，导致数据在安全共享、交换和处理时的复杂度大幅提升。

3）个人隐私保护难。健康医疗数据中包含特别敏感的个人隐私信息，必须依法进行管控和保护；对涉及健康医疗数据的管理要以相应的法律法规做指导，在进行健康医疗数据的收集、存储、挖掘等应用时，需要解决个人隐私保护的难题。

4. 电商行业大数据

电商行业作为基于互联网技术衍生的新型业务，积累了大量商家数据、买家数据、商品数据，以及在买卖交易过程中产生的订单数据、交易数据和用户行为数据等。借助大数据技术发展契机，电商行业也开始了大数据时代的转型。电商行业基于长期积累的海量数据，开始在不同业务方向利用大数据技术分析、挖掘数据价值。

（一）精细化运营及管理

通过对电商数据分析和利用，改进电商业务运营模式，实现业务精细化的运营及管理，包括业务分析、业务智能化、精细化营销、风险管理和运营效率等方面。例如在营销方面，能够通过对以往的营销数据分析，最大化的利用数据资源建立适用的营销方案，并及时通过营销反馈数据的收集和分析，反作用于营销方案的改进工作；在风险管理方面，通过建立实

时响应的风险监控系统，对电商业务流中的数据与风险大数据进行关联分析，更好地识别和控制业务风险。

（二）提升业务效率

通过对电商行业中的业务数据进行大数据分析，作用于电商企业的产品设计、绩效管理、配送效率、库存管理和客户关系管理等关键环节，提升业务能力和效率。例如，通过对历史销售数据进行大数据分析，结合后期的产业环境，商家可以预测销售数据，进而优化物流和仓储；通过客户大数据分析，物流企业能够更合理的选择派送方式、优选路径，并提供差异化服务，提高物流服务质量等。

（三）改进消费体验

电商企业通过对消费者数据的分析可以产生如消费者“画像”等衍生数据。企业基于这些衍生数据，可以为消费者提供个性化的服务，如个性化商品推荐、个性化搜索以及智能机器人客服等服务，提升消费体验。

（四）保障生态圈良性发展

电商业务的生态圈涉及电商业务平台、商家、消费者、为商家提供服务的独立软件提供商，以及相关服务机构等众多合法参与者，但也存在着诸如诈骗组织和炒信团伙之类的谋求非法利益的黑灰产组织。电商行业可利用大数据技术精准识别风险，打击炒信、欺诈和侵权等恶意行为，促进电商生态的良性发展。

电商行业大数据在促进业务发展的同时，相应的安全挑战也随之浮现，主要表现在：

1）数据权属不清

电商业务的开展主要包括电商平台、商家和消费者三方，电商业务产生的数据如何划分其所有权、控制权和使用权，是在电商业务中合理使用数据的前提。当前电商业务的大数据应用中，通常利用电商平台对数据进行分析，也存在商家或商家授权独立软件提供商使用商家数据进行分析的

情况，在权利归属不明确的情况下，责任的归属也难以界定，相关数据安全难以保障。

2）大数据聚合分析风险

电商业务的大数据应用涉及对消费者相关的数据分析，虽然可以通过隐私保护政策、用户授权协议的形式获取相关数据的使用合法授权，而且在对电商业务分析的过程中也会采用匿名化处理的方式，保证用户的个人信息安全。但是，在对大数据加工计算的过程中，如何保障不会因为大数据的聚合分析而实现“去匿名化”，依然是亟待解决的难题。

3）数据版权保护

电商生态圈内的数据流动和共享较为普遍，目前主要通过法律协议方式约束对数据的使用。但由于缺乏有效的数据版权保护技术手段及措施，难以甄别是否存在超出范围的数据扩散或使用问题。

4）数据跨境安全

目前国家大力支持跨境电商业务，而跨境电商业务必然涉及数据的跨境问题。不同国家和地区的数据保护法规对数据跨境流动的要求存在差异性，比如俄罗斯明确提出俄罗斯公民的数据应在俄罗斯境内更新后方可传到海外进行处理；欧盟则扩大了数据保护法律适用的管辖范围。这些法规将给跨境电商企业带来高昂的合规成本，制约了跨境电子商务的发展。如何处理数据跨境安全合规与跨境电商战略发展的矛盾，是亟待解决的难题。

5. 电信行业大数据

电信运营商拥有大量的数据资源，如网络信息、用户终端信息、用户位置信息等，同时电信行业近年来利用大数据进行深度挖掘分析，将丰富的网络、用户等数据资源加工抽取后封装为服务，向客户提供。可拓展的大数据应用服务主要为内部支撑、社会服务、商业运营等几大类，典型应

用主要包括：

（一）运营商内部支撑

运营商利用大数据技术，改善服务水平，提升用户体验。根据网络流量变化、网络信令数据信息和各个设备长期运行情况，运营商利用大数据分析，可以及时调整资源配置，进行全网络优化，提升网络质量和网络利用率；可以监控服务质量，改进突发故障自动应对机制；还可以通过分析不同用户群流量使用特征以及存量用户流量趋势，按需提供不同的流量服务，实现流量经营。

（二）社会服务支撑

运营商利用大数据技术，对其拥有的各种数据进行深度挖掘，根据不同行业客户的业务特点，提供城市规划与交通线网规划、路网状态实时监测与公共交通调度、信息验真服务、公共区域安全监测等不同的社会服务支撑信息服务，如基于移动蜂窝网络产生的位置信息，为政府公共管理、城市规划、交通规划提供数据，为零售商提供人群分布、流向、热点等信息。

（三）商业运营支撑

运营商利用大数据技术，基于用户终端信息、用户基础数据、订购产品及行为数据等，可以生成用户基础数据特征与兴趣特征模型，进而预测客户行为，进行程序化广告投放，实现精准营销。还可以通过用户画像进行产品与推广的规划，以个性化、精准型的业务内容不断增强客户黏性。

大数据给电信行业带来新的发展机遇，电信运营商借助已有的数据积累优势，不断发展大数据应用，但同时数据的集中管理、数据对外开放等新技术特点和业务新形态应用，也使电信行业大数据面临新的安全风险和挑战，主要包括：

1）供应链安全

通信数据在移动网络设备中产生，而这些设备是由多家供应商提供。

同时，存在大数据平台系统第三方供给代建设、代维护等问题，在特定阶段，部分设备的操作权在供应商手中，这意味着供应链的各环节存在安全风险。

2）数据集中管理

在大数据业务应用发展的驱动下，电信运营商的数据由原来的各系统分散存储转变为大数据平台集中存储模式，大数据资源的安全风险更加集中，一旦发生安全事件将涉及海量客户信息及公司数据资产。

3）平台组件开源

大数据平台多使用开源软件，这些软件设计初衷主要考虑高效数据处理，缺乏安全性保障，滞后于电信业务发展的安全防护能力，存在安全隐患。

4）敏感数据共享

在电信运营商内部信息系统建设相对分散，敏感数据跨部门、跨系统共享留存比较常见，其中一旦存在系统安全防护措施不当，均可能发生敏感数据泄漏，造成“一点突破、全网皆失”的严重后果。

附录B

大数据应用安全实践

我们在推广大数据应用的同时，大数据的安全问题不容忽视。需要推动大数据安全关键技术和大数据安全解决方案的研究。在研究大数据安全技术和制定大数据安全标准时，业界的一些大数据安全实践为我们制定切实可行的大数据安全解决方案，以及制定大数据安全关键标准提供重要参考。

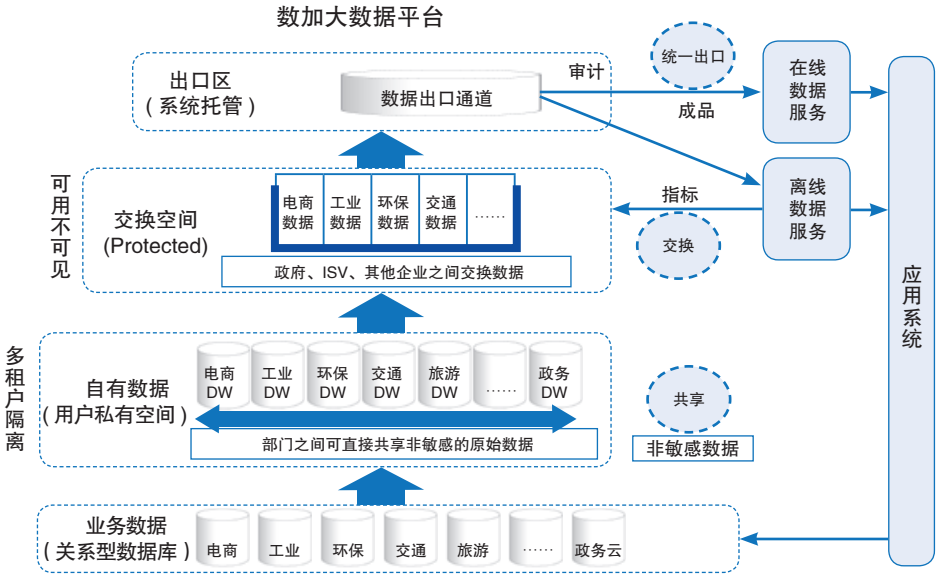
为了解决大数据应用过程中各方关切的数据安全和个人信息保护问题，国内外大数据应用厂商和研究机构纷纷开展大数据安全问题研究，并推出包括加密、访问控制和审计等安全功能的大数据安全解决方案或安全实践。我们收集了大数据开源组织和国内外一部分厂商在大数据安全方面的安全实践[1]。下面将简要介绍这些安全实践。

1. 阿里云大数据安全实践

阿里云数加大数据平台提供从数据采集，加工、数据分析、机器学习到最后数据应用的全链路技术和服务。基于阿里云数加大数据平台，除了可以打造智能可视化透明工厂、智能交通实时预测和实时监控监测、智能医院就医接诊服务，以及大数据网络安全态势感知系统外，还可以打造成一个满足政府不同部门以及政企之间实现数据共享的数据交换平台。为了

[1] 本部分的大数据安全应用实践来自于参与本白皮书撰写单位的贡献。安全实践中所涉及的知识产权属于各贡献单位。本白皮书也不对各贡献单位的安全实践有效性进行评判。

保障数据共享和交换过程中的数据安全，数加大数据平台通过安全机制和管控措施实现不同用户之间数据的“可用不可见”，具体如图B-1所示：



图B–1 阿里云数加大数据交换平台安全框架

为确保数据交换和共享的安全，避免数据滥用，阿里云数加平台提供了一系列安全措施，包括：

1）**密钥管理和鉴权**。提供统一的密钥管理和访问鉴权服务，支持多因素鉴权模型；

2）**访问控制和隔离**。实施多租户访问隔离措施，实施数据安全等级划分，支持基于标签的强制访问控制，提供基于ACL的数据访问授权模型，提供全局数据视图和私有数据视图，提供数据视图的访问控制；

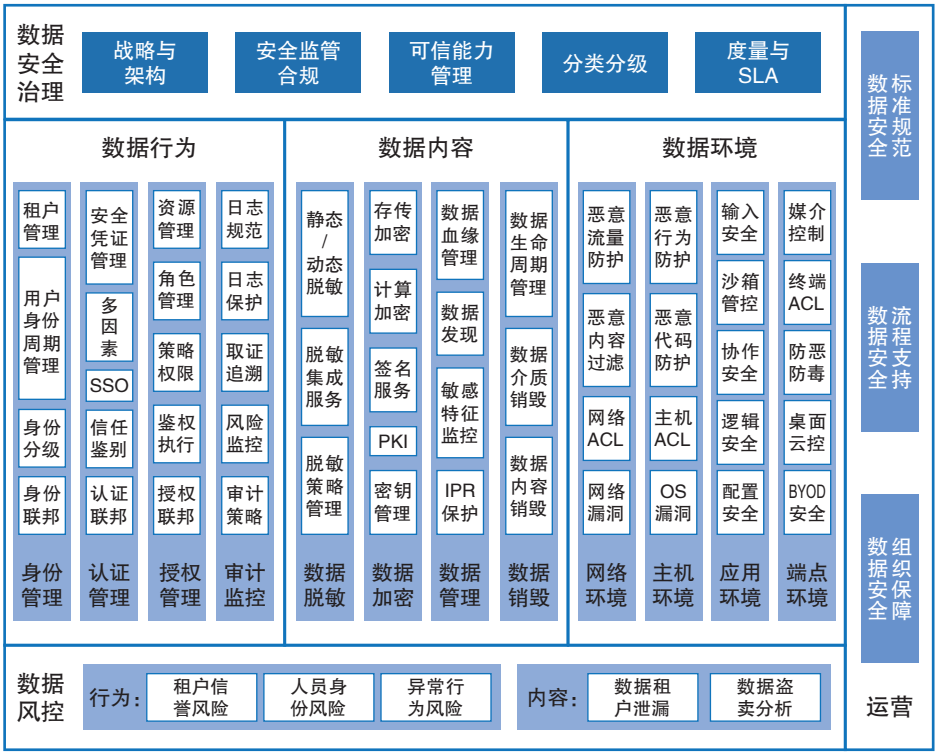
3）**数据安全和个人信息保护**。提供数据脱敏和个人信息去标识化功能，提供满足国产密码算法的用户数据加密服务；

4）**安全审计和血缘追踪**。提供数据访问审计日志，支持数据血缘追踪，跟踪数据的流向和衍生变化过程；

5）**审批和预警**。支持数据导出控制，支持人工审批或系统预警；提供数据质量保障系统，对交换的数据进行数据质量评测和监控、预警；

6）**生命周期管理**。提供从采集、存储、使用、传输、共享、发布、到销毁等基于数据生命周期的技术和管理措施。

阿里云基于数据生命周期构建全面的数据安全保障体系，从数据行为、数据内容、数据环境等角度提供技术和管理措施，具体如图B-2所示：



图B–2 阿里云大数据安全管控体系

通过实施阿里云大数据安全管控体系，提供“可用不可见”的大数据交换共享平台安全环境，以保障大数据在“存储、流通、使用”过程中的安全。

2. 百度大数据安全实践

数据是百度公司的重要资产。百度公司在内部构建了公司级大数据平台，收录公司各个业务领域的数据，建设数据闭环解决方案，推动全公司数据的统一管理、数据共享、数据发现和数据使用。这些聚在一起的数据资产来自多个部门和业务，对安全的要求也不同。百度非常重视大数据应用过程中的安全保障，在安全方面形成了统一的大数据安全框架，通过在数据全生命周期各环节实施安全技术和管理机制，为大数据平台和用户数据提供安全保障。

（一）百度大数据平台安全架构

百度大数据平台具备基础的系统安全、安全管理，以及以数据安全分级机制为核心的数据安全架构，如图B-3所示：

图B–3 百度大数据平台安全架构

系统安全和安全管理是百度大数据平台中最基础的安全机制。数据安全架构在整个大数据安全架构中处于极为重要的位置。数据安全架构包括安全审计、安全控制和安全加密三部分，并采用安全分级机制，分为基础级和可选级。安全基础级别包括安全审计和安全控制两个功能，它是所有在大数据平台的业务数据都会得到的安全基础保障，为大数据平台上的数

据提供生命周期过程中的可审计性和细粒度完整控制功能。可选级别包括数据的加解密功能，支持各种强度的加解密算法。百度大数据平台支持数据的加密存储，考虑到平台每天产生的数据量极其庞大，以及数据运算的效率要求，可以根据数据的业务特点和密级要求来选择不同强度的加密算法。

（二）百度大数据平台关键安全能力

百度提出4A安全体系来构建大数据平台的关键安全能力，主要包括：

1）Account（**账号**）：为每个用户创建唯一的用户账号，并对用户身份进行鉴别，确保数据访问控制和安全审计可以追溯到个人账号。同时，采用基于角色的用户分组管理，将系统管理角色、系统数据建设角色和数据查看角色进行区分。

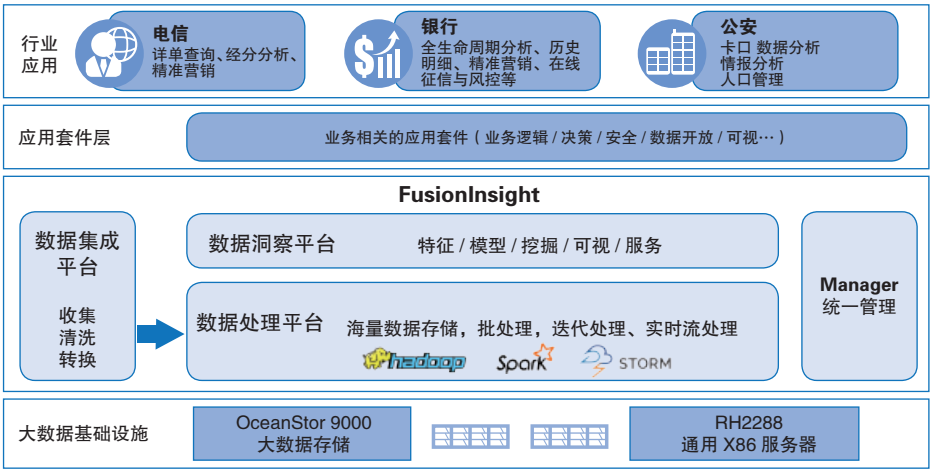
2）Authentication（**鉴别**）：百度大数据平台上的数据访问必须有统一的身份鉴别机制。百度大数据平台采用统一单点登录身份认证技术对用户进行身份鉴别管理。

3）Authorization（**授权**）：百度大数据平台需要根据数据访问主体身份，以及被访问数据的密级，实现对各类数据的访问授权。对于机密等级以上的数据，需要对接到具体的电子审批流程。此外，数据在流转过程中，大数据平台可以自动判断对应的下一个节点的安全等级和人员授权情况，进行数据流转的安全判断和维护。

4）Audit（**审计**）：百度大数据平台具有审计日志记录功能，实现对系统中针对用户管理、权限管理、用户登陆、数据获取/访问/修改等行为的完整日志记录。基于系统审计日志，可以实现事中的安全监控，以及事后的行为溯源和取证分析。

3. 华为大数据安全实践

华为大数据分析平台FusionInsight基于开源社区软件Hadoop进行功能增强，提供企业级大数据存储、查询和分析的统一平台，帮助企业快速构建海量数据信息处理系统。FusionInsight是完全开放的大数据分析平台，并针对金融、运营商等数据密集型行业的运行维护、应用开发等需求打造了高可靠、高安全、易使用的运行维护系统和全量数据建模中间件。华为FusionInsight大数据分析平台框架图如图B-4所示。

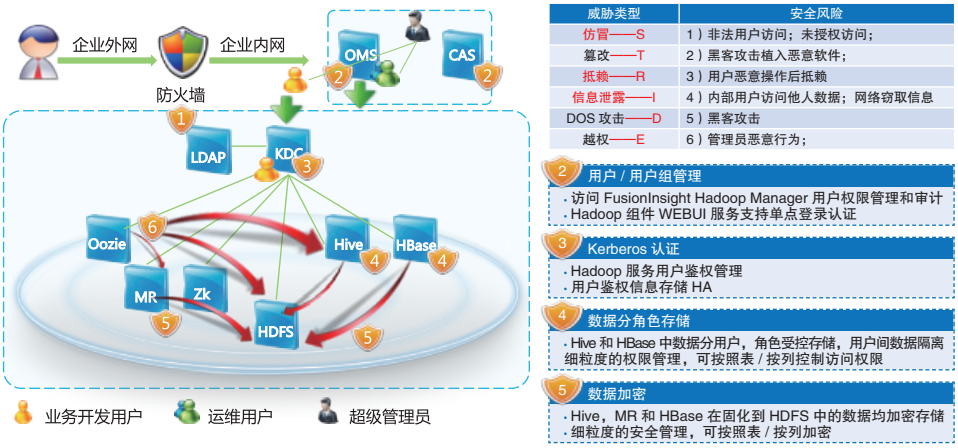


图B–4 华为FuisonInsight大数据分析平台框架图

大数据分析平台汇聚着大量数据，面临着更多的安全威胁和挑战，包括数据滥用和用户隐私泄露问题。华为FuisonInsight大数据分析平台提供可运营的安全体系，从网络安全、主机安全、用户安全和数据安全方面提供全方位的安全防护（如图B-5）：

（一）网络安全：FusionInsight集群支持通过网络平面隔离的方式保证网络安全。

（二）主机安全：通过对FusionInsight集群内节点的操作系统安全加



图B–5 华为FuisonInsight大数据分析平台安全体系图

固等手段保证节点正常运行，包括更新最新补丁、操作系统内核安全加固、操作系统权限控制、端口管理、部署防病毒软件等。

（三）用户安全：通过提供身份认证、权限控制、审计控制等安全措施防止用户假冒、越权、恶意操作等安全威胁：

1）身份认证。FusionInsight使用LDAP作为帐户管理系统，并通过Kerberos对帐户信息进行安全认证，统一了Manager系统用户和组件用户的管理及认证，提供单点登录。

2）权限控制。基于用户和角色的认证统一体系，遵从帐户/角色RBAC（基于角色的访问控制）模型，实现通过角色进行权限管理，对用户进行批量授权管理，降低集群的管理难度；通过角色创建访问组件资源的权限，可以细粒度管理资源（例如文件、目录、表、数据库、列族等访问权限）；将角色授予用户/用户组，简化用户/用户组的权限配置。

3）审计日志。FusionInsight审计日志中记录了用户操作信息，可以快速定位系统是否遭受恶意的操作和攻击，并避免审计日志中记录用户敏感信息：确保每一项用户的破坏性业务操作被记录审计，保证用户业务操作

可回溯；为系统提供审计日志的查询、导出功能，可为用户提供安全事件的事后追溯、定位问题原因及划分事故责任的重要手段。

（四）数据安全：从集群容灾、备份、数据完整性、数据保密性等方面保证用户数据的安全。

1）文件系统加密：Hive、HBase可以对表、字段加密，集群内部用户信息禁止明文存储；

2）加密灵活：加密算法插件化，可进行扩充，亦可自行开发。非敏感数据可不加密，不影响性能；

3）业务透明：上层业务只需指定敏感数据（Hive和HBase表级、列级加密），加解密过程业务完全不感知。

（五）数据容灾：FusionInsight集群容灾为集群内部保存的用户数据提供实时的异地数据容灾功能；它对外提供了基础的运维工具，包含主备集群关系维护，数据重建，数据校验，数据同步进展查看等功能。

4. 京东大数据安全实践

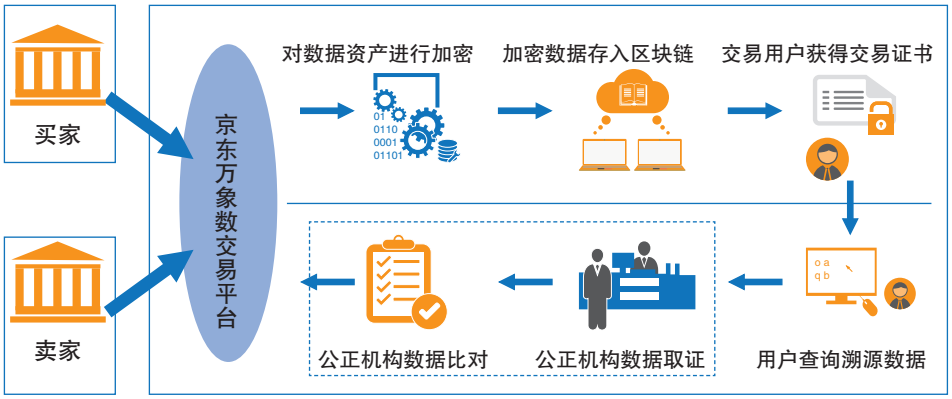
数据资源已经成为一种基础战略资源，数据的共享和流通会产生巨大价值。然而，数据资源在流通过程中却面临着诸多瓶颈和制约，尤其是当数据一种特殊的数字内容产品时，其权益保护难度远大于传统的大数据，一旦发生侵权问题，举证和追责过程都十分困难。

为了解决这些问题，京东万象数据服务平台（如图B-6所示）利用区块链技术对流通的数据进行确权溯源，数据买家在数据服务平台上购买的每一笔交易信息都会在区块链中存储起来，数据买家通过获得交易凭证可以看到该笔交易的数字证书以及该笔交易信息在区块链中的存储地址，待买家需要进行数据确权时，登录用户中心进入查询平台，输入交易凭证中的相关信息，查询到存储在区块链中的该笔交易信息，从而完成交易数据的溯源确权。

京东万象数据服务平台主要通过数据交易平台和区块链溯源平台2个核心模块提供服务：

1）数据交易平台。平台通过数据搜索、数据展示、数据评论等服务，以各种维度展示数据商品，并提供订单和支付系统完成用户数据交易；

2）区块链溯源平台。用户订单信息、数据标识、交易私钥等交易信息存入区块链集群中，用户获得交易凭证，并可利用该溯源平台查询溯源数据，完成数据资产确权。



图B–6 京东万象数据服务平台数据安全框架

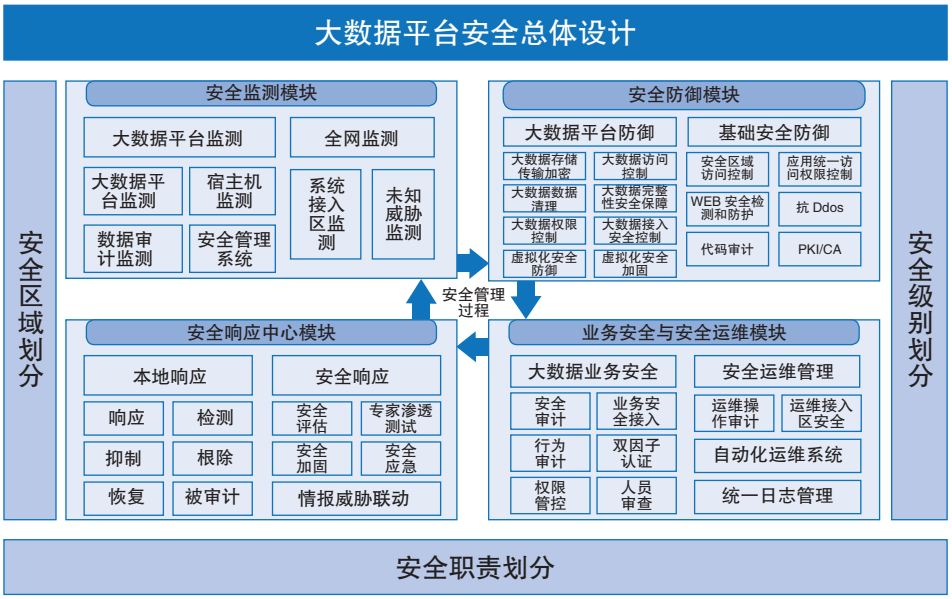
在安全保障方面，为了防止数据流通过程中的个人身份冒用问题，京东万象数据服务平台通过使用公安部提供的个人身份认证服务对用户身份进行识别和保护。京东万象数据服务平台结合公安部eID技术，该技术密码技术为基础、以智能安全芯片为载体、由“公安部公民网络身份识别系统”签发给公民的网络身份标识，能够在不泄露身份信息的前提下在线远程识别用户身份。京东万象数据服务平台通过区块链溯源和eID技术，有效解决了合法用户基于互联网开展大数据安全交易的数字产品版权保护问题，保障了数据拥有者在数据交易中的合法权益。

5. 奇虎360大数据安全实践

奇虎360在面对日益严峻的安全挑战时，不断更新技术思路，实现了及时响应最新的网络安全威胁。为应对千变万化的网络安全威胁，奇虎360通过部署的数万台大数据服务器，对当前网络安全事件进行实时监测与分析，采用大数据技术对网络安全威胁进行跟踪和防范。

为了保障安全，大数据平台依照“安全三同步”原则进行建设，即同步规划、同步组织实施、同步运作投产。

奇虎360的大数据平台安全保障体系框架如图B-7所示。大数据平台安全保障体系框架包括“安全职责划分”，“安全区域划分”，“安全级别划分”，“安全监测模块”，“安全防御模块”，“业务安全与安全运维模块”，“安全响应中心模块”等部分。



图B–7 奇虎360大数据平台安全保障体系框架

1)安全职责划分。安全职责划分是整体方案的基础，所有技术手段都应贴近安全职责划分，为其服务。梳理大数据平台各方安全责任边界，对整个活动中的安全事件进行详细的责任划分。

2)安全区域划分。大数据平台环境相对复杂，涉及多类业务，多类系统，现有网络结构已经考虑了分级问题，在此基础上，需进一步细化安全域的划分以及不同安全域、不同安全级别的访问控制设计。

3)安全级别划分。按照安全区域划分结果，为每个区域制定响应的安全等级，区域安全等级与用户安全等级、数据安全等级相互对应。通过安全级别的划分确保可信合规使用资源。

4)安全监测模块。其中主要包括大数据平台安全防御审查系统并提供基于人工或自动化的多层次的安全监测服务。

5)安全防御模块。按照统一规划、统一标准的设计思路，在充分考虑当前网络应用和实际环境的基础上，对整体的网络划分为若干个安全域和安全区，建设大数据平台面向各个区域的基础安全防御系统和大数据平台自身的防御系统。

6)业务安全与安全运维模块。实现安全运维操作的分级管理，针对大数据业务安全和安全运维工作的用户赋予符合其安全职责划分的权限，实现业务安全和安全运维。

7)安全响应中心模块。采用本地响应+安全响应的新型工作模式。本地响应实现当前问题的及时规范化处理，安全响应结合云端的情报威胁联动、本地终端协调联动、以及专家等提供及时的技术保障服务。

6. 腾讯大数据安全实践

腾讯一直把大数据应用作为公司的重要发展战略，并依托十多年的互联网产品开发和运营经验，形成了一套完整、可靠、扩展性强的大数据业务应用框架，为用户提供大数据处理服务。腾讯大数据业务应用框架为用

户提供三大基础能力：

1）**数据**：提供海量的数据接入能力与处理能力；

2）**连接**：提供开放接口，做互联网+的连接器；

3）**安全**：重视网络安全，将其作为连接一切的防护体系。

腾讯特别注重在提供大数据处理服务过程中的数据安全和隐私保护问题，采取安全技术和管理措施确保大数据业务的健康发展。大数据和云计算密不可分，腾讯云通过端、主机、网络、业务的安全服务，为客户提供安全的大数据业务。腾讯大数据安全涉及的安全关注重点如图B-8所示。

<table>
<tr><th></th><th>相关系统</th><th>使用前审查</th><th>使用中控制</th><th>使用后审计</th></tr>
<tr><td>管理安全</td><td>管理系统</td><td rowspan="3">认证、授权、
授信管理</td><td>分级管理</td><td rowspan="3">审计管理</td></tr>
<tr><td>传输安全</td><td>输入输出系统
处理系统</td><td>接口安全
中间层安全
…</td></tr>
<tr><td>数据安全</td><td>处理系统
框架系统</td><td>存储安全
抹除安全
…</td></tr>
<tr><td>平台安全</td><td>框架系统</td><td>权限管理</td><td>系统防御</td><td>操作审计</td></tr>
</table>

图B–8 腾讯公司大数据安全关键点

1）**平台安全**。关注系统自身的安全性，防止来自系统层面的攻击，同时为更高级安全防御措施提供系统级别的支持，包括：系统防御，即防御来自系统层面的攻击，如漏洞攻击、嗅探攻击、流量攻击（如DDoS）等；权限管理，即提供文件、设备等底层资源的权限管理能力，防止越权访问；操作审计：即提供文件、设备等底层资源的访问、操作历史日志，为更高级的审计提供数据和功能支持。

2）**数据安全**。关注数据生命周期各阶段的安全性，防止数据丢失、覆盖、篡改带来的损失。包括：存储安全，即采用多副本方式存储数据，

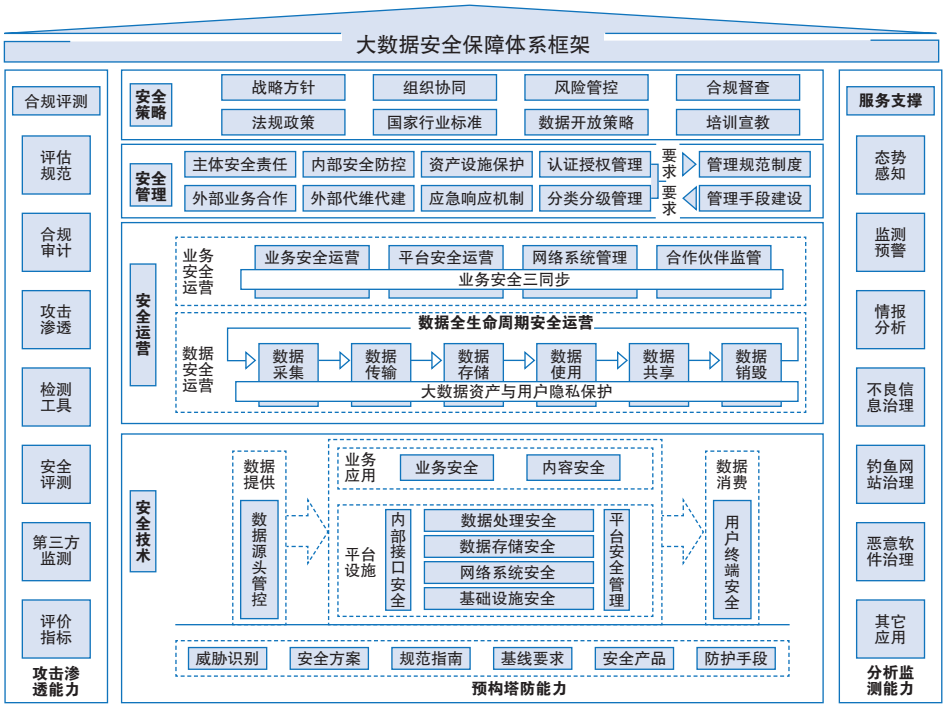
防止数据非正常丢失；抹除安全，即数据延迟删除，防止误操作带来的数据丢失。

3）传输安全。关注数据在传输过程中的安全性，包括：接口安全，即采用安全接口设计及高安全的数据传输协议，保证在通过接口访问、处理、传输数据时的安全性，避免数据被非法访问、窃听或旁路嗅探；中间层安全，即使用加密等方法隐藏实际数据，保证数据在通过中间层的过程中不被恶意截获，只有数据管理者通过密钥等方式可以在平台中动态解密并访问原始数据。

4）安全管理。关注对大数据分析平台的合理、合规使用，通过与技术配套的管理手段控制风险，保证安全。包括：认证、鉴权、授信管理，即确保用户对平台、接口、操作、资源、数据等都具有相应的访问权限，避免越权访问；分级管理，即根据敏感度对数据进行分级，对不同级别的数据提供差异化的流程、权限、审批要求等管理措施，数据安全等级越高，管理越严格；审计管理，基于底层提供的审计数据，在权限管理、数据使用、操作行为等多个维度上对大数据分析平台的运转提供安全审计能力，确保及时发现大数据分析平台中的隐患点，视不同严重程度采取包括排除隐患、挽回数据、人员追责在内的多种补救措施，同时指导大数据分析平台不再重复类似的问题。

7. 中国移动大数据安全实践

为应对大数据应用服务过程中数据滥用和个人隐私安全风险，中国移动建立了完善的大数据安全保障体系，目标是保护大数据权属性、保密性、完整性、可用性、可追溯性，实现大数据“可管、可控、可信”，保护公司各领域大数据资产及用户隐私。大数据安全保障体系框架如图B-9所示。



图B–9 中国移动大数据安全保障体系框架图

中国移动大数据安全保障体系涉及安全策略、安全管理、安全运营、安全技术、合规评测、服务支撑等六大体系：

1）**安全策略体系**是在遵循国家大数据安全政策框架的基础上，开展顶层设计，明确公司大数据安全总体策略，指导相关管理制度、技术防护、安全运营、合规评测、服务支撑工作的开展，是其它体系建设的基本依据。

2）**安全管理体系**是通过管理制度建设，明确运营方安全主体责任，落实安全管理措施，相关制度包括第三方合作管理、内部安全管理、数据分类分级管理、应急响应机制、资产设施保护和认证授权管理等安全管理规范要求。

3）**安全运营体系**是通过定义运营角色，明确运营机构安全职责，实

现对大数据业务及数据的全流程、全周期安全管理，通过对大数据的平台系统、业务服务、数据资产和用户隐私的有效安全运营管控，保障业务可持续健康发展。

4）**安全技术体系**建设目标是有效预构塔防能力，包括基础设施、网络系统、数据存储、数据处理以及业务应用等层次安全防护。通过制定涉及网络、平台、系统、数据、业务系列安全技术规范支撑开展安全防护能力建设。

5）**安全合规评测体系**建设目标是持续优化安全评估能力，通过合规评估、安全测试、攻击渗透等手段，实现对大数据业务各环节风险点的全面评估，保障安全管理制度及技术要求的有效落实。

6）**大数据服务支撑体系**的理念是“安全保数据、数据促安全”，重点是基于大数据资源为信息安全保障提供支撑服务，如基础安全态势感知、数据安全监测预警、情报分析舆情监测、以及不良信息治理等安全领域的应用。通过开展大数据在大数据安全管控等各个领域的应用研究，为信息安全管控提供新型的支撑服务手段。

同时，中国移动对用户个人信息的各个处理环节施行严格规定与落实：

1）对客户信息所包含的内容进行界定、分类及分级；

2）明确信息安全管理责任部门及职责。对各部门的职责进行了严格要求和细致规定，并明确相关岗位角色及权限；

3）对客户敏感信息操作进行严格管理。对于涉及用户敏感信息的关键操作，严格遵守金库模式保护要求，采取“关键操作、多人完成、分权制衡”的原则，实现操作与授权分离；

4）设立客户信息安全检查制度；

5）不断提高客户信息系统技术管控水平；

6）严控第三方信息安全风险。

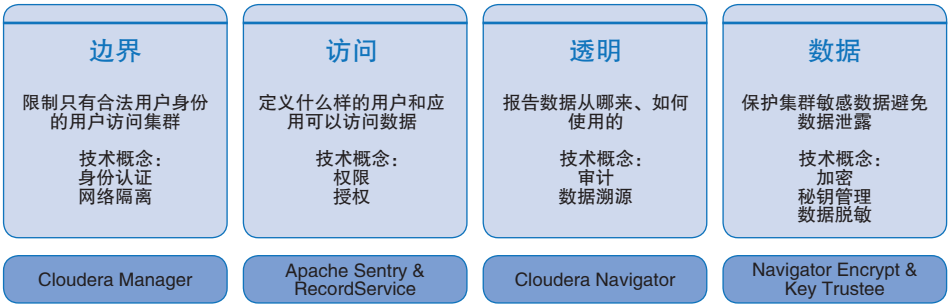
另外，中国移动自主研发了大数据安全管理平台——雷池，实现数据的统一认证、集中细粒度授权、审计监控、数据脱敏以及异常行为检测告警，可对数据进行全方位安全管控，做到事前可管、事中可控、事后可查。

8. Cloudera大数据安全实践

Hadoop已经广泛应用于金融、电信、制造、能源以及健康医疗领域，这些领域的客户基于Hadoop搭建企业数据湖，完成企业数据整合。数据整合之前是存放在相对独立的系统进行安全存储及管理。数据整合之后，原本只有少数人访问到的数据分享给更多的用户进行分析，如何有效的对访问者进行身份审核，数据的权限管理，数据访问留痕即审计，以及对涉密程度比较高的数据在大数据平台进行加密，是企业数据湖面临的重要问题。

Cloudera在大数据安全保障方面，提供了从数据平台身份认证、访问授权管理、数据加密保护到安全审计全流程的安全解决方案体系架构。Cloudera大数据平台安全体系架构如图B-10所示：

1）边界。关注于控制外部用户或者服务对集群的访问过程中的身份鉴别，也称之为身份认证模块，这是实施大数据安全架构的基础；在



图B–10 Cloudera安全体系架构

Cloudera数据平台中所有组件都能提供基于Kerberos的认证功能，某些组件还能提供额外的基于LDAP（Active Directory）或者是SAML的认证；用户在访问启用了安全认证的集群时，必须能通过服务所需要的安全认证方式。在部署身份认证时，根据的企业基础设施不同，可以选择不同的部署解决方案。

2）访问。关注于用户或者应用访问数据时，对用户的权限定义和实施过程，通常称为授权；Cloudera可以限定用户是否有对某种资源的访问能力。基于Hadoop的数据平台通常都提供了多样化的资源和服务，但受限于访问控制措施，不得不限制了Hadoop使用的广度和深度。起初Hadoop仅仅是作为ETL的补充开放给SQL开发者使用，后来各业务分析部门意识到Hadoop的便利性，也需要相应数据和服务的访问授权，这就要求大数据平台需要和企业现有LDAP或者AD进行整合，同时能给不同应用提供一致的基于角色的访问控制能力。Cloudera通过Apache Sentry来完成对大数据系统访问策略的配置和权限控制实施，从而可以实现一致的访问权限控制配置和实施过程，比如说，一个用户通过Hive或者Impala对某张表实施了权限配置，那么此用户通过Spark或者Search访问这个数据时，Apache Sentry同样能确保一致的权限控制效果。

3）透明。理解数据的来源，以及知道数据怎么被使用的，对监测大数据系统中是否存在非法数据访问非常关键，这需要通过安全审计来实现。安全审计的目的是捕获系统内的完整活动记录，且不可被更改。Navigator提供了自动化的数据上下游关系收集，并能进行可视化展示。对任何一个Hadoop上的数据源，细致到数据表的一个列，可以抽取这个列是由上游的哪些数据源、哪些列，生成了下游数据源的哪些列。

4）数据。提供数据在传输过程及静态存储的加密保护，在敏感数据被越权访问时仍然能够得到有效保护。 Cloudera推荐通过Cloudera Manager配置TLS来完成数据在传输过程的加密，数据的静态加密可以

通过HDFS Data-at-Rest Encryption，Navigator Encrypt以及Navigator Key Trustee来完成。关于加密的秘钥管理，Cloudera平台除了支持传统基于Java KeyStore的加密密钥管理方式外，还提供了Navigator Key Trustee服务提供更好的秘钥存储方案，它还能提供和企业现有的HSM（Hardware Security Module）集成解决方案。

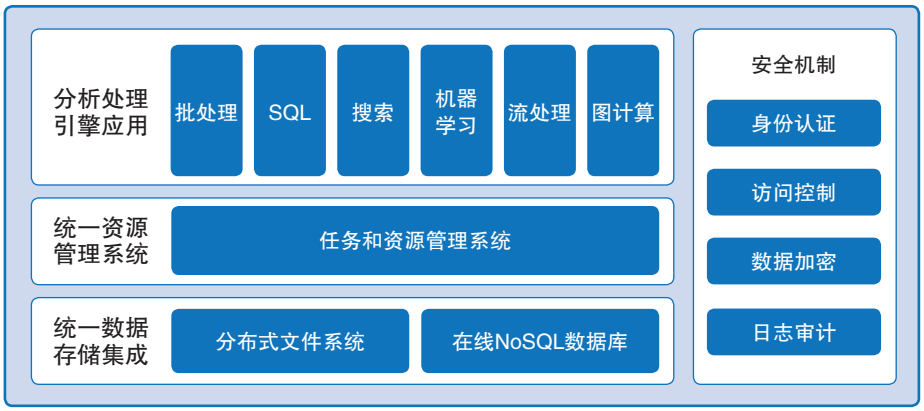
通过Cloudera Manager提供的向导式操作界面，方便启用Hadoop的Kerberos认证，避免企业用户受到黑客勒索攻击。Sentry为大数据平台的组件Hive，Impala，Solr以及HDFS提供细粒度的基于角色的权限管理功能，避免数据集中后的非授权访问。Navigator提供大数据平台所有组件的统一审计功能。Navigator Encrypt保障数据传输过程及静态存储都是以加密形式存在避免黑客截取数据及数据泄露。与此同时，Cloudera也在不断加强Hadoop生态系统的安全特性，比如RecordService为Hadoop平台提供统一的安全管控。增强Kudu，Spark等技术在数据存储及处理的安全。

9. Hadoop大数据安全实践

当前，以Hadoop为基础的大数据开源生态圈应用非常广泛。最早，Hadoop考虑只在可信环境内部署使用，而随着越来越多部门和用户加入进来，任何用户都可以访问和删除数据，从而使数据面临巨大的安全风险。另外，对于内部网络环境和数据销毁过程管控的疏漏，在大数据背景下，如不采取相应的安全控制措施，也极易出现重大的数据泄露事故。

为了应对上述安全挑战，2009年开始，Hadoop开源社区开始注重保护大数据安全，相继加入了身份验证、访问控制、数据加密和日志审计等重要安全功能，如图B-11所示。

身份验证是确认访问者身份的过程，是数据访问控制的基础。在身份验证方面，Hadoop大数据开源软件将Kerberos作为目前唯一可选的强安全的认证方式，并以此为基础构建安全的大数据访问控制环境。基于身



图B–11 开源大数据平台安全机制

份验证的结果，Hadoop使用各种访问控制机制在不同的系统层次对数据访问进行控制。HDFS（Hadoop分布式文件系统）提供了POSIX权限和访问控制列表两种方式，Hive（数据仓库）则提供了基于角色的访问控制，HBase(分布式数据库)提供了访问控制列表和基于标签的访问控制。数据加密作为保护数据安全、避免数据泄漏的主要手段在大数据应用系统中广泛采用，有效地防止通过网络嗅探或物理存储介质销毁不当而导致数据泄密。对于数据传输，Hadoop对各种数据传输提供了加密选项，包括对客户端和服务进程之间以及各服务进程之间的数据传输进行加密。同时Hadoop也提供了数据在存储层落盘加密，保证数据以加密形式存储在硬盘上。最后，Hadoop生态系统各组件都提供日志和审计文件记录数据访问，为追踪数据流向，优化数据过程，以及发现违规数据操作提供原始依据。

基于上述系列安全机制，Hadoop基本构建起了满足基本安全功能需求的大数据开源环境。Kerberos作为事实上的强安全认证方式被业界广泛采用。但由于Kerberos采用对称密钥算法来实现双向认证，在大规模部署基于Kerberos的分布式认证系统时，可能会带来部署和管理上的挑战。普遍解决方案是采用第三方提供的工具简化部署和管理流程。访问控制方面，大数据环境访问控制的复杂性不仅在于访问控制的形式多样，另一方

在于大数据系统允许在不同系统层面广泛共享数据，需要实现一种集中统一的访问控制从而简化控制策略和部署。数据加密方面，通过基于硬件的加密方案，可以大幅提高数据加解密的性能，实现最低性能损耗的端到端和存储层加密。然而，加密的有效使用需要安全灵活的密钥管理，这方面开源方案还比较薄弱，需要借助商业化的密钥管理产品。日志审计作为数据管理，数据溯源以及攻击检测的重要措施不可或缺。然而Hadoop等开源系统只提供基本的日志和审计记录，存储在各个集群节点上。如果要对日志和审计记录做集中管理和分析，仍然需要依靠第三方工具。

10. IBM 大数据安全实践

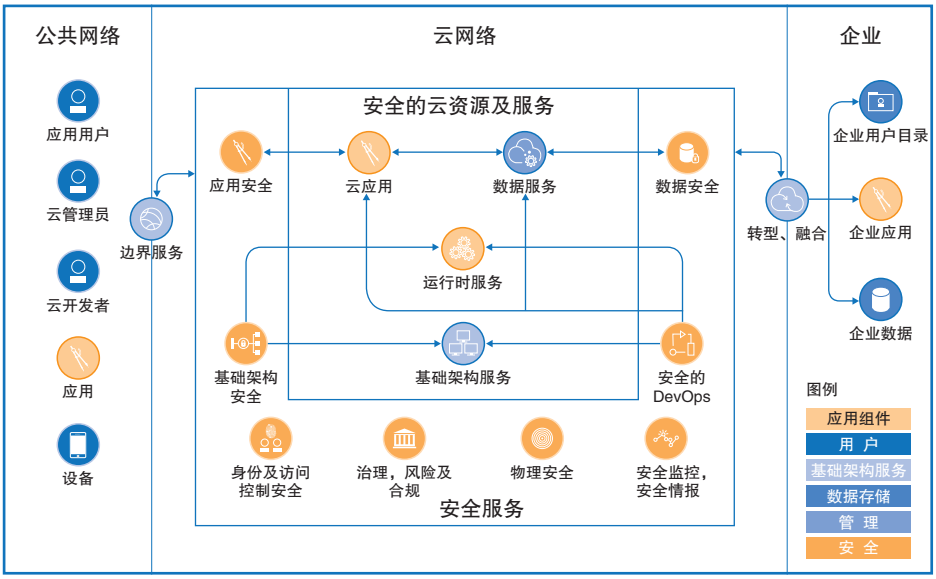
IBM Security Guardium是一个完整的数据安全平台，提供了一套完整的能力，比如敏感数据的发现和分类、分级，安全性评价，数据和文件活动检测，通过伪装，阻断，报警和隔离保护敏感数据。Guardium不仅保护数据库，它还被扩展到保护数据仓库、ECM、文件系统和大数据环境(Hadoop or NoSQL)等。除了安全平台，IBM架构提供了云上应用构建的实践。IBM为大数据分析和安全开发了客户云架构，这个构架作为参考架构和行业标准在CSCC（Cloud Standards Customer Council）发布，它描述了使用云计算托管大数据分析解决方案的厂商中立的最佳实践及构成这个架构的所有组件的细节。这个参考架构的所有组件都可以用开源技术实现。

（一）IBM安全参考架构和数据安全

如图B-12所示，IBM安全参考架构提供了保护云上部署，开发和运维的安全组件的概览。

在谈及数据安全时我们通常需要区分静态数据和动态数据。数据安全旨在发现、分类和保护云数据和信息资产，重点在于对静态数据和动态数据的保护。IBM数据安全架构包括所有数据类型，如传统企业数据及大数据环境中任意形式的数据（结构化的和非结构化的）。IBM数据安全架构

囊括了基于治理、风险和合规的数据安全所需要的各个模块，以下总结了云计算解决方案中需要考量的数据安全相关的关键模块。



图B–12 IBM安全参考架构

1）数据保护

一个完备的云计算数据保护解决方案需要考虑将以下服务选项提供给客户：

- 云环境中的静态数据加密
- 存储块和文件存储加密服务
- 使用IBM Cleversafe的对象存储加密
- 使用 IBM Cloud Data Encryption Services(ICDES)的数据加密服务
- 基于云的硬件安全模块 (HSM)
- 使用IBM Key Project的密钥管理和证书管理

针对以上的每一个服务选项，都需要制定一套具体的流程、控制方案和实施策略用于实施。

2）数据完整性

数据完整性旨在维护和保证数据在其整个生命周期中的准确性和一致性。在本文的语境中，数据完整性指的是如何防范数据被外界篡改。数据的哈希值可用于检测数据是否被非法篡改。这个方法可以用于对静态数据和动态数据提供保护。

3）数据分类和数据活动监测

数据分类是帮助保护关键信息安全的有效方法。在保护敏感信息之前，必须确定和鉴别它的存在。自动化发现和分类过程，是防止泄漏敏感信息数据保护策略的关键组件。 Guardium 提供了集成的数据分类能力和无缝的方法，来发现、鉴别和保护最关键数据，不管是在云上还是在数据中心。

Guardium 也可以提供数据活动监测，以及通过认知分析来发现针对敏感数据的异常活动，防止未授权的数据访问，也提供可疑活动的警报，自动化合规性流程，并抵御内部和外部攻击。

4）数据隐私和法律法规

数据隐私决定了在相关政策和法律法规所规定的范围内，如何对信息（特别是与个人相关的信息）进行采集、使用、分享和处置。

根据IBM的政策，每一个云服务都需要实现技术上和组织上的安全和隐私保护措施。这些措施都是根据云服务的架构、使用目的及服务类型来实现的。无论服务的类型，IBM关于每一个云服务的具体管理责任，都会在相关的协议中列出。

5）IBM大数据智能安全

IBM大数据智能安全，合并了IBM QRadar智能安全平台的实时的安全关联和异常发现能力以及法庭取证的能力，和由BigInsights 提供的包括定制的大规模结构化数据（比如安全设备告警，操作系统日志，DNS事务和网络流）和非结构化数据（比如电子邮件，社交内容，数据包信息和业务

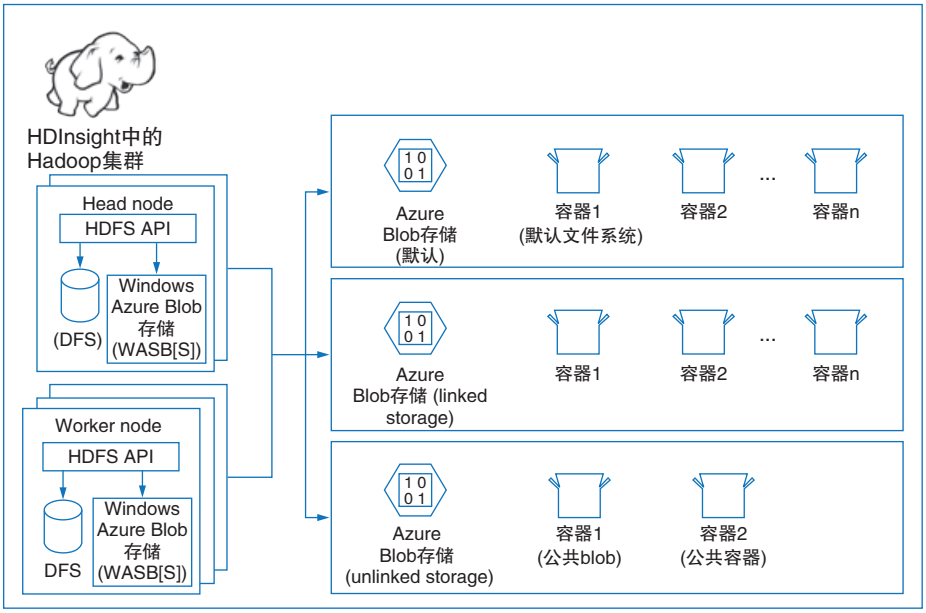
交易）的分析和发现能力。

11. Microsoft大数据安全实践

HDInsight是微软运行在Micsoroft Azure上的大数据服务。Azure HDInsight 以云方式部署并设置 Apache Hadoop 集群，从而提供旨在对大数据进行管理、分析和报告的软件框架。微软的大数据服务Azure HDInsight支持多种数据技术，包括基本的Hadoop 分布式文件系统HDFS，超大型表格的非关系型数据库HBase，类似SQL的查询Hive，分布式处理和资源管理MapReduce和YARN等等，如图B-13所示：

HDInsight作为Azure云服务的一部分，Azure从多个方面提供了安全保护，其中包括：

1）**使用 Azure Blob 存储**。Azure Blob存储是一种与Hadoop兼容的选项，是一种稳健、通用的存储解决方案，它与HDInsight无缝集成。通过



图B–13 HDInsight 存储体系结构的抽象视图

Hadoop分布式的文件系统HDFS界面，可以针对Blob存储中的结构化或非结构化数据直接运行 HDInsight 中的整套组件。通过将数据存储在Blob存储中，可以安全删除用于计算的HDInsight集群而不会丢失用户数据。

2）**密钥保管库**。安全的密钥管理对在云中保护数据必不可少。借助Azure密钥保管库，可以通过使用硬件安全模块 (HSM)中存储的密钥对密钥和小密文密码进行加密。为了增加保障，可以在 HSM 中导入或生成密钥。如果选择这样做，Microsoft将使用FIPS 140-2 第2级认证的 HSM处理用户的密钥。密钥保管库设计用于确保Microsoft不会看到或提取用户的密钥。通过Azure日志记录监视并审核密钥的使用情况——将日志传送到Azure HDInsight或SIEM中以进行额外的分析和威胁检测。

3）**多重身份验证**。Azure 多重身份验证是要求使用多种方式（而不仅仅是用户名和密码）对用户的身份进行验证的一种方法。它为用户登录和事务提供了附加的安全层。Azure多重身份验证可帮助保护对数据和应用程序的访问，同时可以满足用户对简单登录过程的需求。它通过各种简单的验证选项（例如电话、短信、移动应用通知或验证码）来提供强大的身份验证。

4）Azure Active Directory(Azure AD)。Azure AD是 Microsoft 提供的基于多租户云的目录和标识管理服务。Azure AD 包含整套标识管理功能，例如多重身份验证、设备注册、自助密码管理、自助组管理、特权帐户管理、基于角色的访问控制、应用程序使用情况监视、多样化审核以及安全监视和警报。

通信地址：北京市东城区安定门东大街1号
中国电子技术标准化研究院（100007）
联 系 人： 叶润国　金涛
联系电话：010– 64102735