ICS 35.040
L 80

中华人民共和国国家标准

GB/T 35274—2017

信息安全技术
大数据服务安全能力要求

Information security technology—
Security capability requirements for big data services

2017-12-29 发布　　2018-07-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会　发布

目　　次

前　言

本标准按照 GB/T 1.1—2009 给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由全国信息安全标准化技术委员会(SAC/TC 260)提出并归口。

本标准起草单位:清华大学、中国电子技术标准化研究院、中国信息安全测评中心、阿里巴巴(北京)软件服务有限公司、中国移动通信集团公司、浙江蚂蚁小微金融服务集团有限公司、阿里云计算有限公司、启明星辰信息技术集团股份有限公司、联想(北京)有限公司、四川大学、工业和信息化部计算机与微电子发展研究中心(中国软件评测中心)、华为技术有限公司、中国电子科技网络信息安全有限公司、深圳市腾讯计算机系统有限公司、中电长城网际系统应用有限公司、陕西省信息化工程研究院、广州赛宝认证中心服务有限公司、天津南大通用数据技术股份有限公司、西安未来国际信息股份有限公司、深信服科技股份有限公司、中国科学院信息工程研究所(信息安全国家重点实验室)、中国科学院软件研究所、北京京东叁佰陆拾度电子商务有限公司、国家信息技术安全研究中心、北京匡恩网络科技有限责任公司、腾讯云计算(北京)有限责任公司、北京奇虎科技有限公司、北京数聚世界信息技术有限公司、西北大学。

本标准主要起草人:叶晓俊、叶润国、谢安明、王建民、刘贤刚、陈兴蜀、胡影、陈星、陈雪秀、李克鹏、江为强、闵京华、张勇、王禹、周波、孙茵茵、程广明、黄少青、任兰芳、王永霞、葛小宇、望娅露、落红卫、梅婧婷、赵伟、李汝鑫、金涛、刘璘、郭晓雷、马红霞、刘玉岭、张辉文、刘伯仲、李小丁、都婧、代威、陈锦、任望、孙骞、张滨、冯运波、罗永刚、鲍旭华、朱红儒、周润松、孙彦。

引　言

大数据服务是针对数量巨大、种类多样、流动速度快、特征多变等特性的数据集，通过底层可伸缩的大数据平台和上层多种大数据应用，提供覆盖数据生命周期相关数据活动的一种网络信息服务。大数据服务提供者要确保大数据平台与应用安全可靠地运行，满足保密性、完整性、可用性等大数据服务安全目标。

本标准将大数据服务安全能力分为一般要求和增强要求两个级别。一般要求是指大数据服务提供者在开展大数据服务时，能够抵御或应对常见的威胁，能将大数据服务受到破坏后的损失控制在有限的范围和程度内，具备基本的事件追溯能力。增强要求是指在大数据服务涉及国家安全，或对经济发展和社会公共利益有较大影响时，大数据服务提供者具备一定的主动识别并防范潜在攻击的能力，能高效应对安全事件并将其损失控制在较小范围内，能保证安全事件追溯的有效性、大数据服务的可靠性、可扩展性和可伸缩性。根据所承载数据的重要性和大数据服务不能正常提供服务或遭受到破坏时可能造成的影响范围和严重程度，大数据服务提供者的安全能力要求也各不相同。

信息安全技术
大数据服务安全能力要求

1 范围

本标准规定了大数据服务提供者应具有的组织相关基础安全能力和数据生命周期相关的数据服务安全能力。

本标准适用于对政府部门和企事业单位建设大数据服务安全能力，也适用于第三方机构对大数据服务提供者的大数据服务安全能力进行审查和评估。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。

GB/T 22239—2008 信息安全技术 信息系统安全等级保护基本要求

GB/T 25069—2010 信息安全技术 术语

GB/T 31168—2014 信息安全技术 云计算服务安全能力要求

GB/T 35273—2017 信息安全技术 个人信息安全规范

GB/T 35295—2017 信息技术 大数据 术语

3 术语和定义

GB/T 25069—2010 和 GB/T 35295—2017 界定的以及下列术语和定义适用于本文件。

3.1

大数据 big data

具有数量巨大、种类多样、流动速度快、特征多变等特性，并且难以用传统数据体系结构和数据处理技术进行有效组织、存储、计算、分析和管理的数据集。

3.2

数据生命周期 data lifecycle

数据从产生，经过数据采集、数据传输、数据存储、数据处理(包括计算、分析、可视化等)、数据交换，直至数据销毁等各种生存形态的演变过程。

3.3

数据服务 data service

提供数据采集、数据传输、数据存储、数据处理(包括计算、分析、可视化等)、数据交换、数据销毁等数据生存形态演变的一种网络信息服务。

3.4

大数据服务 big data service

支撑机构或个人对大数据采集、存储、使用和数据价值发现等数据生命周期相关的各种数据服务和系统服务。

注：大数据服务一般面对的是海量、异构、快速变化的结构化、半结构化和非结构化数据服务，且通过底层可伸缩的

大数据平台和上层各种大数据应用的系统服务提供。

3.5

大数据应用　big data application

执行数据生命周期相关的数据采集、数据传输、数据存储、数据处理(如计算、分析、可视化等)、数据交换、数据销毁等数据活动,运行在大数据平台,并提供大数据服务的各种应用系统。

3.6

大数据平台　big data platform

采用分布式存储和计算技术,提供大数据的访问和处理,支持大数据应用安全高效运行的软硬件集合,包括监视大数据的存储、输入/输出、操作控制等大数据服务软硬件基础设施。

3.7

大数据服务提供者　big data service provider

通过大数据平台和应用,提供大数据服务的机构。

3.8

大数据使用者　big data consumer

使用大数据平台或应用的末端用户、其他信息技术系统或智能感知设备。

3.9

大数据系统　big data system

包括大数据使用者、大数据服务提供者、大数据应用和大数据平台的信息系统。

3.10

数据供应链　data supply chain

对大数据服务提供者的数据采集、数据预处理、数据聚合、数据交换、数据访问等相关数据活动进行计划、协调、操作、控制和优化所需的可用数据资源形成的链状结构。

注:数据供应链目标是将大数据服务所需的各种数据和系统资产,通过计划、协调、操作、控制、优化等数据活动,确保大数据服务提供者能在正确的时间,按照正确的数据服务协议送给正确的大数据使用者。

3.11

数据交换　data interchange

为满足不同平台或应用间数据资源的传送和处理需要,依据一定的原则,采取相应的技术,实现不同平台和应用间数据资源的流动过程。

3.12

数据共享　data sharing

让不同大数据用户能够访问大数据服务整合的各种数据资源,并通过大数据服务或数据交换技术对这些数据资源进行相关的计算、分析、可视化等处理。

3.13

重要数据　important data

我国机构和个人在境内收集、产生的不涉及国家秘密,但与国家安全、经济发展以及公共利益密切相关的数据。

注:重要数据通常指公共通信和信息服务、能源、交通、水利、金融、公共服务、电子政务等重要行业和领域的各类机构在开展业务活动中收集和产生的,不涉及国家秘密,但一旦泄露、篡改或滥用将会对国家安全、经济发展和社会公共利益造成不利影响的数据(包括原始数据和衍生数据)。

4　概述

4.1　总体要求

大数据服务提供者应依据 GB/T 31168—2014 和 GB/T 22239—2008,从信息技术(Information

Technology，简称 IT）角度对大数据服务基础设施采取必要的安全管控措施，保障大数据平台和应用的系统服务安全可靠地运行和大数据服务的业务使命。本标准只规定了通过大数据平台和应用，提供大数据服务的机构应具备的基础安全要求和数据服务安全要求：

a） 基础安全要求：大数据服务提供者要创建大数据服务安全策略和规程，建立系统和数据资产清单、组织和人员岗位，通过对大数据服务的安全规划和需求分析形成满足大数据服务的元数据结构、符合业务流程的数据供应链结构和数据服务接口规范，以及符合法律法规和相关标准要求的大数据服务基础安全能力要求。

b） 数据服务安全要求：大数据服务提供者要针对数据生命周期相关的数据活动，形成数据采集、数据传输、数据存储、数据处理、数据交换、数据销毁等数据服务安全要求，以降低大数据服务中数据生命周期安全管理相关的安全风险，保障大数据服务的业务使命和数据安全。

大数据服务提供者应依据大数据服务的数据保护价值、大数据服务类型，结合自身的大数据服务模式、大数据服务角色、大数据服务目标和支撑大数据服务的基础设施（参见附录 A），选择本标准列举的大数据服务安全能力要求项进行建设和评估。

注：大数据服务涉及的数据资源可能依赖于其他机构的数据服务或系统服务，则大数据服务提供者以合同、协议或其他方式对供应链上各参与方的相应安全责任进行规定并予以落实，要求他们具备与大数据服务提供者相当的安全防护能力。

4.2 要求分级

本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根据大数据系统所承载数据重要性和大数据服务异常可能造成的影响范围和严重程度，遵循如下保护要求：

a） 大数据服务中没有承载重要数据，且该服务不能正常提供服务或遭受破坏时，对国家经济发展和社会公共利益影响有限，对国家安全和国计民生没有影响的，按照一般要求进行安全保护；

b） 大数据服务中承载有重要数据，或在不能正常提供服务或遭受破坏时，对国家经济发展和社会公共利益造成的影响较大，或者会影响国家安全和国计民生的，按照增强要求进行安全保护。

5 基础安全要求

5.1 策略与规程

5.1.1 一般要求

大数据服务提供者应：

a） 制定符合机构大数据服务战略规划的安全策略，明确安全方针、安全目标和安全原则。

b） 确保所制定的安全策略覆盖数据生命周期相关的数据服务和系统服务，内容应包括目的、范围、岗位、责任、管理层承诺、内外部协调及合规性要求等。

c） 制定安全策略相关的规程，并将安全策略和规程分发至机构大数据服务部门、岗位和人员。

d） 制定并实施与安全策略和规程相适应的大数据平台和应用安全实施细则，包括外部数据资源整合、数据共享、数据发布等数据供应链安全管理细则、合同要求及审核机制。

e） 定期审核和更新大数据服务安全策略和规程。

5.1.2 增强要求

大数据服务提供者应：

a） 在组织架构发生重大调整或大数据服务业务发生重大变化时，及时评估安全策略与规程的实施效果，并将效果反馈到安全策略和规程文件的修订过程中。

b） 对大数据服务安全策略和规程进行体系化的评估，制定大数据服务安全能力提升计划。

5.2 数据与系统资产

5.2.1 数据资产

5.2.1.1 一般要求

大数据服务提供者应：

a） 建立数据资产安全管理规范，明确大数据服务相关数据资产的安全管理目标和安全原则。

b） 建立数据资产分类分级方法和操作指南，以及数据资产分类分级的变更审批流程和机制。

c） 建立数据资产组织和管理模式，制定数据资产登记制度，明确数据资产管理相关方。

d） 建立数据资产清单，明确大数据服务相关数据资产管理范围和属性。

e） 定期审核和更新数据资产安全管理相关的安全规范、管理制度等。

5.2.1.2 增强要求

大数据服务提供者应：

a） 依据数据资产和数据主体安全分级要求建立相应的标记策略、访问控制、数据加解密、数据脱敏等安全机制和管控措施。

b） 建立大数据服务所需的机构内外部数据资产的安全治理原则和数据资源整合规范。

c） 建立机构级数据资产管理平台，实现对数据资产的统一管理。

5.2.2 系统资产

5.2.2.1 一般要求

大数据服务提供者应：

a） 建立大数据系统资产安全管理规范，明确系统资产安全管理目标和安全原则。

b） 建立大数据系统资产建设和运营管理制度和机制，明确规划、设计、采购、开发、运行、维护及报废等资产管理过程的安全要求。

c） 建立大数据系统资产登记机制，形成大数据服务的系统软硬件资产清单，明确系统资产安全责任主体及相关方，并及时更新系统资产相关信息。

d） 建立大数据系统资产分类和标记规程，使资产标记易于填写和依附在相应的系统资产上。

e） 定期审核和更新大数据系统资产管理相关的安全规范、管理制度。

5.2.2.2 增强要求

大数据服务提供者应：

a） 建立大数据系统资产管理平台，具备系统资产统一注册、管理和使用监控等能力。

b） 建立大数据系统资产更新、运营风险评估和 IT 系统供应链安全审查规程和制度。

5.3 组织和人员管理

5.3.1 组织管理

5.3.1.1 一般要求

大数据服务提供者应：

a) 建立大数据服务安全管理组织机构，明确大数据服务安全岗位和岗位用户职责。
b) 建立大数据服务安全领导小组，指定机构最高管理者或授权代表担任小组组长，并明确组长责任与权力。
c) 指定大数据服务平台与应用安全规划、安全建设、安全运营和系统维护工作的责任部门。
d) 建立机构内部监督管理职能部门，对大数据服务和各用户操作行为进行安全监督管理。
e) 制定大数据服务安全追责制度，定期对责任部门和安全岗位组织安全检查，形成检查报告。

5.3.1.2 增强要求

大数据服务提供者应：

a) 建立体系化的大数据服务安全管理机构，机构最高管理人员应作为大数据服务安全领导小组组长，且配备必要的管理人员和技术人员。
b) 设置专职的大数据服务安全岗位，建立规范化的大数据服务安全保护、评估及考核专职队伍。

5.3.2 人员管理

5.3.2.1 一般要求

大数据服务提供者应：

a) 制定大数据服务人力资源安全策略，明确不同岗位人员在数据生命周期各阶段数据服务和系统服务相关的工作范畴和安全管控措施。
b) 制定大数据服务人员招聘、录用、上岗、调岗、离岗、考核、选拔等人员安全管理制度。
c) 在录用重要岗位人员前对其进行背景调查，确保符合相关的法律、法规、合同和道德要求，并与所有涉及大数据服务岗位人员签订安全责任协议。
d) 明确大数据服务重要岗位的兼职和轮岗、权限分离、多人共管等安全管理要求。
e) 建立人员安全责任奖惩管理制度，并按照规定对造成大数据服务损失的人员给予相应的处理，记录并保存相关信息。
f) 建立第三方人员安全管理制度，对接触个人信息、重要数据等数据的人员进行审批和登记，并要求签署保密协议，定期对这些人员行为进行安全审查。
g) 在重要岗位人员调离或终止劳动合同时，与其签订保密协议。

5.3.2.2 增强要求

大数据服务提供者应：

a) 明确关键岗位人员背景调查范围，定期对关键岗位人员进行背景审查。
b) 明确关键岗位人员安全能力要求，并确定他们培训技能考核内容与考核指标，定期对关键岗位人员进行审查和能力考核。

5.3.3 角色管理

5.3.3.1 一般要求

大数据服务提供者应：

a) 建立大数据服务相关的安全角色，明确安全角色的分配策略和授权范围。
b) 建立用户角色及角色权限冲突的定期审查机制，及时更新用户角色及角色权限授权信息。
c) 明确大数据服务相关重要岗位及其角色安全要求，建立重要岗位角色清单和授权机制。

5.3.3.2 增强要求

大数据服务提供者应：

a) 依照大数据系统架构建立分层的角色体系、职责分离等大数据服务安全角色管理机制。

b) 建立用户应用上下文感知的角色启动、停用与禁用的动态管理策略、规程和机制。

5.3.4 人员培训

5.3.4.1 一般要求

大数据服务提供者应：

a) 制定大数据服务安全岗位人员的安全培训计划，并对培训计划定期审核和更新。

b) 制定关键岗位转岗、岗位升级等相应的人员安全培训计划，并对培训计划定期审核和更新。

c) 按计划对相关人员开展安全培训，包括政策、法律、法规、标准等合规性培训，并对培训结果进行评价、记录和归档。

5.3.4.2 增强要求

大数据服务提供者应根据不同安全岗位要求，开展大数据服务安全实际操作技能培训与考核。

5.4 服务规划与管理

5.4.1 战略规划

5.4.1.1 一般要求

大数据服务提供者应：

a) 明确机构大数据服务类型、大数据平台与应用战略规划目标，编制大数据服务战略规划，确保其安全规划符合网络安全法等国家法律法规及网络安全等级保护等制度相关要求。

b) 依据机构战略规划目标，制定大数据服务安全规划各阶段目标、任务和工作重点，并对战略规划目标和安全规划实施过程进行监督与控制。

c) 成立战略规划评估小组，负责机构安全规划评估，确保大数据服务安全策略、安全目标和战略规划内容的合规性。

5.4.1.2 增强要求

大数据服务提供者应：

a) 建立大数据服务安全规划动态调整制度，并通过信息化平台进行管理。

b) 建立大数据服务所涉及数据安全管理纲领性文件，包括但不限于：数据治理、数据质量、元数据，以及平台与应用安全相关的数据所有权、数据开放与共享等安全策略。

5.4.2 需求分析

5.4.2.1 一般要求

大数据服务提供者应：

a) 建立大数据服务安全需求分析的制定流程和评审制度，明确安全需求文档内容要求。

b) 依据国家法律、法规、标准与主管机构的政策规范要求，分析大数据服务安全合规性需求。

c) 结合机构战略规划、大数据服务业务目标和业务特点，分析大数据服务业务安全需求。

d) 识别大数据服务面临的威胁和自身脆弱性，分析大数据服务安全风险和应对措施需求。

e) 依据机构战略规划，明确大数据服务安全需求和安全规划实施的优先级。

5.4.2.2 增强要求

大数据服务提供者应使用数据驱动分析等方法进行大数据服务安全需求分析，确保大数据服务相关安全需求的有效制定和规范化表达。

5.4.3 元数据安全

5.4.3.1 一般要求

大数据服务提供者应：

a) 建立大数据服务相关元数据及其管理规范，如数据域、字段类型、表结构、逻辑存储和物理存储结构及管理方式。
b) 建立大数据服务安全架构相应的安全元数据管理规范，如口令策略、权限列表、授权策略。
c) 建立元数据访问控制策略，明确元数据管理角色及其授权控制机制。
d) 建立元数据操作的审计制度，确保元数据操作的可追溯。

5.4.3.2 增强要求

大数据服务提供者应：

a) 具备实现大数据服务元数据统一管理的能力。
b) 依据资产分类分级策略建立元数据安全属性自动分级机制。
c) 依据元数据安全属性建立标记策略及标记定义和标记管理机制。

5.5 数据供应链管理

5.5.1 数据供应链

5.5.1.1 一般要求

大数据服务提供者应：

a) 建立数据供应链安全管理规范和安全方针，明确数据供应链安全目标、原则和范围。
b) 确保数据供应链上下游对数据交换、使用和利用符合法律法规，并有技术保障措施。
c) 明确数据供应链上下游责任和义务，确保数据供应链相关数据服务真实可用。
d) 通过合作协议方式明确大数据服务数据供应链中数据的使用目的、供应方式、保密约定等。
e) 建立数据供应链目录和相关数据源数据字典，明确数据供应链的责任部门和人员。
f) 对数据供应链上下游的大数据服务提供者和大数据使用者的行为进行合规性审核和分析。

5.5.1.2 增强要求

大数据服务提供者应定期对数据供应链上下游数据活动安全风险和数据安全管理能力进行评估。

5.5.2 数据服务接口

5.5.2.1 一般要求

大数据服务提供者应：

a) 制定数据服务接口安全控制策略，明确规定使用服务接口的安全限制和安全控制措施，如身份鉴别、授权策略、访问控制机制、签名、时间戳、安全协议等。
b) 明确服务接口安全规范，包括接口名称、接口参数、接口安全要求等，具备对接口不安全输入参数进行限制或过滤能力，为接口提供异常处理能力。

c) 具备服务接口访问的审计能力，并能为大数据安全审计提供可配置的数据服务接口。
d) 对大数据平台与应用内跨安全域间的接口调用采用安全通道、加密传输等安全机制。

5.5.2.2 增强要求

大数据服务提供者应建立服务接口安全监管措施，以对接口访问进行必要的自动化监控和处理。

5.6 合规性管理

5.6.1 个人信息保护

5.6.1.1 一般要求

大数据服务提供者应依据 GB/T 35273—2017 相关要求，采取必要的技术手段或管控措施，建立符合国家法律法规和相关标准的个人信息保护能力。

5.6.1.2 增强要求

大数据服务提供者应：

a) 在大数据应用及关联业务组件下线以及设备退网时，要妥善转移、转存、销毁保存的个人信息，避免因人工管理模式改变或机构业务重组与兼并等方式而规避个人信息保护要求。
b) 具备对大数据服务中个人信息处理操作行为进行溯源和合规性分析的能力。
c) 建立针对多源数据集汇聚和关联后个人信息利用的安全风险分析和保护控制措施。
d) 具备评价大数据服务中的个人信息去标识有效性的能力。

5.6.2 重要数据保护

5.6.2.1 一般要求

大数据服务提供者应：

a) 建立符合网络安全法等法律法规的重要数据的安全策略、规范、制度和管控措施。
b) 在数据生命周期相关数据服务中涉及重要数据时要确保满足法律法规及相关标准要求。
c) 建立管控措施和采取技术手段控制重要数据流向，避免因人工管理模式改变或机构业务重组和兼并等方式而规避重要数据保护要求。
d) 建立重要数据监控机制，具备对重要数据生命周期相关操作行为进行合规性分析的能力。
e) 定期对重要数据安全策略、规范、制度和管控措施进行风险评估，并及时对其进行调整和更新。

5.6.2.2 增强要求

大数据服务提供者应：

a) 提供重要数据的自动化脱敏机制与措施，支持如匿名、泛化、随机、加密等脱敏方法。
b) 建立大数据服务中数据沉淀防控机制，评估和防范使用沉淀数据获得重要数据的风险。
c) 具备评价大数据服务中重要数据脱敏有效性的评估能力。

5.6.3 数据跨境传输

5.6.3.1 一般要求

大数据服务提供者应：

a) 制定数据跨境传输业务处理流程和数据跨境传输审批制度，明确数据跨境传输安全策略、管理制度、管理规范和管控措施。
b) 按照我国个人信息和重要数据出境安全评估办法、数据出境安全评估指南等国家法律、法规和

标准对数据跨境传输业务进行安全评估，确保数据跨境传输的合法性和正当性。

c) 具备对大数据服务中的数据跨境传输处理操作行为进行合规性分析的能力。

5.6.3.2 增强要求

大数据服务提供者应定期或在发生重大信息安全事件后，对数据跨境传输有关制度、流程和技术进行审查和检验，记录审查和检验结果并提交机构最高级大数据服务安全管理组织审批。

5.6.4 密码支持

5.6.4.1 一般要求

大数据服务提供者应按照国家密码管理规定使用和管理有关密码技术和设施，并按规定生成、分发、存取、更新、备份和销毁密钥。

5.6.4.2 增强要求

大数据服务提供者应：

a) 具备密钥集成管理的能力，并满足密钥管理互操作性等有关标准规范。

b) 具备密文数据透明处理能力。

6 数据服务安全要求

6.1 数据采集

6.1.1 数据分类分级

6.1.1.1 一般要求

大数据服务提供者应：

a) 按照数据资产分类分级策略对采集数据进行分类分级标识。

b) 对不同类别和级别的采集数据实施相应的安全管理策略和保障措施。

c) 具备对数据分类分级变更操作进行合规性审核的能力。

6.1.1.2 增强要求

大数据服务提供者应依据数据分类分级策略变更对相关历史数据进行归档，并记录数据分类分级变更过程，确保数据分类分级过程的可追溯性。

6.1.2 数据收集和获取

6.1.2.1 一般要求

大数据服务提供者应：

a) 制定数据采集原则，明确采集数据的目的和用途，确保数据收集和获取的合法性和正当性。

b) 明确数据收集和获取源、数据收集范围和频度，确保数据收集和获取仅限数据业务所需的数据，且是与其大数据服务相关。

c) 制定数据收集和获取操作规程，规范数据收集和获取渠道、数据格式、流程和方式。

d) 对数据收集和获取环境（如采集渠道）、设施和技术采取必要的安全管控措施，确保采集数据的完整性、一致性和真实性。

e) 明确数据收集和获取过程中个人信息和重要数据的知悉范围和安全管控措施，确保采集数据的合规性、完整性和真实性。

f) 采取必要的技术手段和管理措施保证数据收集和获取过程中个人信息和重要数据不被泄露。

6.1.2.2 增强要求

大数据服务提供者应：

a) 采取必要的技术手段或管控措施，对收集和获取到的数据进行完整性和一致性校验。

b) 跟踪和记录数据收集和获取过程，支持对数据收集和获取操作过程的可追溯性。

6.1.3 数据清洗、转换与加载

6.1.3.1 一般要求

大数据服务提供者应：

a) 制定数据清洗、转换和加载操作相关的安全管理规范，确保清洗和转换前后数据间映射关系。

b) 采取必要的技术手段和管理措施，确保在数据清洗、转换和加载过程中对数据进行保护。

c) 记录并保存数据清洗、转换和加载过程中个人信息、重要数据等数据的处理过程。

6.1.3.2 增强要求

大数据服务提供者应：

a) 采取必要的技术手段和管理措施，在个人信息、重要数据等数据有恢复需求时，保证数据清洗、转换和加载过程中产生问题时能有效的还原和恢复数据。

b) 具备数据清洗、转换和加载数据一致性检测及故障处理能力。

6.1.4 质量监控

6.1.4.1 一般要求

大数据服务提供者应：

a) 建立数据采集过程中质量监控规则，明确数据质量监控范围及监控方式。

b) 明确采集数据质量要素，建立异常事件处理流程和操作规范，指定处理对应质量监控项的责任部门或人员。

c) 定义数据源质量评价要素，制定数据采集质量管控措施的策略和标准。

6.1.4.2 增强要求

大数据服务提供者应：

a) 制定数据质量分级标准，明确不同级别和类型的数据采集、清洗、转换、加载等数据采集处理流程质量要求。

b) 定期对数据质量进行分析、预判和盘点，明确数据质量问题定位和修复时间要求。

6.2 数据传输

6.2.1 一般要求

大数据服务提供者应：

a) 区分安全域内、安全域间等不同的大数据服务相关的数据传输场景，建立相应的数据传输安全策略和规程。

b) 采用满足数据传输安全策略相应的安全控制措施，如安全通道、可信通道、数据加密等。

c) 建立数据传输接口安全管理工作规范，包括安全域内、安全域间等数据传输接口规范。

d) 具备在构建传输通道前对两端主体身份进行鉴别和认证的能力。

e) 具备对传输数据的完整性进行检测的能力以及相应的恢复控制措施。

f) 建立机制对数据传输安全策略的变更进行审核和监控，包括对通道安全配置、密码算法配置、密钥管理等保护措施的审核及监控。

6.2.2 增强要求

大数据服务提供者应建立数据传输链路冗余机制，保证数据传输可靠性和网络传输服务可用性。

6.3 数据存储

6.3.1 存储架构

6.3.1.1 一般要求

大数据服务提供者应：

a) 建立开放可伸缩数据存储架构，以满足数据量持续增长、数据分类分级存储等需求。

b) 制定数据存储架构相关的管理规范和安全规则，包括访问控制规则、存储转移安全规则、存储完整性和多副本一致性管理规则等。

c) 采用必要的技术和管控措施保证数据存储架构安全管理规则的实施，确保数据存储完整性和多副本一致性真实有效。

d) 确保存储架构具备对个人信息、重要数据等加密存储能力。

e) 确保存储架构具备数据存储跨机柜或跨机房容错部署能力。

6.3.1.2 增强要求

大数据服务提供者应：

a) 确保存储架构具备数据存储跨地域的容灾能力。

b) 建立满足应用层、数据平台层、操作系统层、数据存储层等不同层次的数据存储加密需求的数据存储加密架构。

6.3.2 逻辑存储

6.3.2.1 一般要求

大数据服务提供者应：

a) 建立数据逻辑存储管理安全规范和机制，以满足不同数据类型、不同数据容量和不同业务需求的逻辑存储安全管理要求。

b) 建立数据分片和分布式存储安全规范和规则，以满足分布式存储下分片数据完整性、一致性和保密性保护要求。

c) 明确数据逻辑存储隔离授权与操作规范、确保具备多租户数据存储安全隔离能力。

6.3.2.2 增强要求

大数据服务提供者应建立分层的逻辑存储授权管理规则和授权操作规范，具备对数据逻辑存储结构的分层和分级保护能力。

6.3.3 访问控制

6.3.3.1 一般要求

大数据服务提供者应：

a) 建立存储系统安全管理员的身份标识与鉴别策略、权限分配策略及相关操作规程。

b) 利用存储访问控制模块实施大数据用户身份标识与鉴别策略、数据访问控制策略、数据扩容及复制策略等，并实现相关安全控制措施。
c) 具备数据分布式存储访问安全审计能力，建立受保护的审计信息存储机制和管控措施。
d) 建立面向大数据应用的安全控制机制，包括访问控制时效的管理和验证，以及应用接入数据存储的合法性和安全性取证机制。

6.3.3.2 增强要求

大数据服务提供者应建立数据存储安全主动防御机制或措施，如基于用户行为或设备行为安全控制机制。

6.3.4 数据副本

6.3.4.1 一般要求

大数据服务提供者应：
a) 建立数据存储冗余策略和管理制度，以满足大数据服务可靠性、可用性等数据安全保护目标。
b) 建立数据冗余强一致性、弱一致性等控制策略与规范，以满足不同一致性水平需求的数据副本多样性和多变性存储管理要求。
c) 建立数据复制、备份与恢复操作过程规范，包括复制、备份和恢复的日志记录规范。
d) 建立数据复制、数据备份与恢复的定期检查和更新工作程序，包括数据副本更新频率、保存期限等，确保数据副本或备份数据的有效性。

6.3.4.2 增强要求

大数据服务提供者应具备数据副本或数据备份存储的多种压缩策略和实现机制，并确保压缩数据副本或数据备份的完整性和可用性。

6.3.5 数据归档

6.3.5.1 一般要求

大数据服务提供者应：
a) 依据数据生命周期和业务规范建立不同阶段数据归档存储相关的操作规程。
b) 建立在线/离线的多级数据归档架构，支持海量数据的有效归档、恢复和使用。
c) 建立归档数据的安全策略和管控措施，确保非授权用户不能访问归档数据。
d) 建立归档数据的压缩或加密策略，确保归档数据存储空间的有效利用和安全访问。
e) 定期地采取必要的技术手段和管控措施查验归档数据完整性和可用性。

6.3.5.2 增强要求

大数据服务提供者应建立归档数据安全审计与恢复制度，并指定专人负责。

6.3.6 数据时效性

6.3.6.1 一般要求

大数据服务提供者应：
a) 制定数据存储时效性管理策略和规程，确保按照法律规定和监管部门的技术规范对相关数据予以记录和保存。
b) 明确数据分享、存储、使用和清除的有效期，具备数据存储时效性授权与控制能力。
c) 建立过期存储数据的安全保护机制，对超出有效期的存储数据应具备再次获取数据控制者授

权的能力。

d) 建立过期存储数据及其备份数据彻底删除方法和机制，能够验证数据已被完全消除或使其无法恢复，并告知数据控制者和大数据使用者。

6.3.6.2 增强要求

大数据服务提供者应：

a) 具备数据时效性自动检测能力，包括但不限于告警、自动清除以及拒绝访问。

b) 为不同时效性的数据建立分层的数据存储方法，具备按照时效性自动迁移数据分层存储的能力，确保大数据用户能高效地获得有效数据。

6.4 数据处理

6.4.1 分布式处理安全

6.4.1.1 一般要求

大数据服务提供者应：

a) 建立分布式处理节点间可信连接策略和规范，采用节点认证等机制来确保节点接入的真实性。

b) 建立分布式处理节点和用户安全属性的周期性确认机制，确保预定义分布式安全策略一致性。

c) 建立分布式处理过程中数据文件鉴别和访问用户身份认证的策略和规范，确保分布式处理数据文件的可访问性。

d) 建立分布式处理过程中不同数据副本节点的更新检测机制，确保这些节点数据拷贝的完整性、一致性和真实性。

e) 建立分布式处理过程中数据泄露控制规范和机制，防止数据处理过程中的调试信息、日志记录、不受控制输出等泄露受保护的个人信息、重要数据等敏感信息。

6.4.1.2 增强要求

大数据服务提供者应：

a) 建立分布式处理外部服务组件注册与使用审核机制。

b) 建立数据分布式处理节点的服务组件自动维护策略和管控措施，包括虚假节点监测、故障用户节点确认和自动修复的技术机制，避免云环境或虚拟环境下潜在的安全攻击。

6.4.2 数据分析安全

6.4.2.1 一般要求

大数据服务提供者应：

a) 建立数据分析相关数据源获取规范和使用机制，明确数据获取方式、访问接口、授权机制、数据使用等。

b) 建立多源数据派生、聚合、关联分析等数据分析过程中的数据资源操作规范和实施指南。

c) 建立数据分析结果输出的安全审查机制和授权控制机制，并采取必要的技术手段和管控措施保证共享数据分析结果不泄露个人信息、重要数据等敏感信息。

d) 对数据分析结果共享的风险进行合规性评估，避免分析结果输出中包含可恢复的个人信息、重要数据等数据和结构标识，如用户鉴别信息的重要标识和数据结构。

e) 对数据分析过程个人信息、重要数据等敏感数据操作进行记录，以备对分析结果质量和真实性进行数据溯源。

6.4.2.2 增强要求

大数据服务提供者应具备基于机器学习的重要数据自动识别、数据安全分析算法设计等数据分析算法及其安全性分析能力，包括外部分析服务组件集成能力。

6.4.3 数据正当使用

6.4.3.1 一般要求

大数据服务提供者应：

a) 确保数据使用和分析处理的目的和范围符合网络安全法等国家相关法律法规要求。

b) 建立数据使用正当性的内部责任制度，保证在数据使用声明的目的和范围内对受保护的个人信息、重要数据等数据进行使用和分析处理。

c) 依据数据使用目的建立相应强度或粒度的访问控制机制，限定用户可访问数据范围。

d) 具备完整的数据使用操作记录和管理能力，以备潜在违约数据使用者责任的识别和追责。

6.4.3.2 增强要求

大数据服务提供者应：

a) 具备信息化技术手段或机制，对数据滥用行为进行有效的识别、监控和预警。

b) 具备违约责任、缔约过失责任、侵权责任等数据使用风险分析和处理能力。

6.4.4 密文数据处理

6.4.4.1 一般要求

大数据服务提供者应建立适合大数据服务特点的数据加密和解密处理策略和密钥管理规范。

6.4.4.2 增强要求

大数据服务提供者应具备对密文数据进行搜索、排序、计算等透明处理的能力。

6.4.5 数据脱敏处理

6.4.5.1 一般要求

大数据服务提供者应：

a) 建立数据脱敏管理规范和制度，明确数据脱敏规则、脱敏方法和使用限制。

b) 明确数据脱敏处理应用场景、数据脱敏处理流程、涉及部门及人员的职责分工。

c) 配置数据脱敏服务组件或技术手段，支持如泛化、抑制、干扰等数据脱敏技术。

d) 能够在屏蔽信息时保留其原始数据格式和特定属性，以满足基于脱敏数据的开发与测试要求。

e) 对数据脱敏处理过程相应的操作进行记录，以满足数据脱敏处理安全审计要求。

6.4.5.2 增强要求

大数据服务提供者应：

a) 明确列出需要脱敏的数据资产，给出不同分类分级数据的脱敏处理流程。

b) 配置脱敏数据识别和脱敏效果验证服务组件或技术手段，确保数据脱敏的有效性和合规性。

c) 明确脱敏数据治理原则和规范。

d) 配置基于策略的数据脱敏支持服务组件或管控措施。

6.4.6 数据溯源

6.4.6.1 一般要求

大数据服务提供者应：

a) 制定数据溯源策略和溯源机制，以及溯源数据安全存储与使用的管理制度。

b) 制定溯源数据表达方式和格式规范，以规范化组织、存储和管理溯源数据。

c) 采用必要的技术手段和管控措施实现分布式数据处理环境下溯源数据采集和存储，确保溯源数据能重现数据处理过程，如追溯操作发起者及发起时间。

d) 对关键溯源数据进行备份，并采取技术手段对溯源数据进行安全保护。

6.4.6.2 增强要求

大数据服务提供者应：

a) 采取技术机制和管控措施保证溯源数据的完整性和保密性。

b) 建立基于溯源数据的数据业务与法律法规合规性审核机制，并依据审核结果增强或改进数据服务相关的访问控制与合规性保障机制和策略。

6.5 数据交换

6.5.1 数据导入导出安全

6.5.1.1 一般要求

大数据服务提供者应：

a) 综合数据量、增长速度、业务需求、性能等因素制定数据导入导出策略与规程。

b) 依据数据分类分级要求建立符合业务规则的数据导入导出安全相关的授权策略、不一致处理策略和流程控制策略。

c) 依据数据导入导出策略与规程、授权策略等，建立数据导入导出安全评估机制和授权审批流程。

d) 对导入导出终端、用户或服务组件等执行身份鉴别，验证其身份的真实性和合法性。

e) 建立存放导出数据介质的标识规范，包括命名规则、标识属性等重要信息，定期验证导出数据的完整性和可用性。

f) 制定导入导出审计策略和审计日志管理规范，并保存导入导出过程中的出错数据处理记录。

g) 采取数据加密、访问控制等技术措施，保障导入导出数据在传输中的保密性、完整性和可用性。

h) 在导入导出完成后对数据导入导出通道缓存的数据进行清除且保证不能被恢复。

6.5.1.2 增强要求

大数据服务提供者应：

a) 采取多因素鉴别技术对数据导入导出操作员进行身份鉴别。

b) 为数据导入导出通道提供冗余备份能力，确保数据安全可靠导入导出要求。

c) 对数据导入导出接口进行流量过载监控，确保海量数据导入导出过程安全可控。

6.5.2 数据共享安全

6.5.2.1 一般要求

大数据服务提供者应：

a) 明确数据共享内容范围和数据共享的管控措施。
b) 明确大数据服务提供者与共享数据使用者的数据保护责任，确保共享数据使用者具备与大数据服务提供者足够或相当的安全防护能力。
c) 明确数据共享涉及机构或部门相关用户职责和权限，保证数据共享安全策略有效性。
d) 审核共享数据应用场景，确保没有超出大数据服务提供者的数据所有权和授权使用范围。
e) 审核共享数据的数据内容，确认属于满足大数据共享业务场景需求范围。
f) 采用数据加密、安全通道等措施保护数据共享过程中的个人信息、重要数据等敏感信息。
g) 制定数据共享审计策略和审计日志管理规范，审计记录详细完整，为数据共享安全事件的处置、应急响应和事后调查提供帮助。
h) 对共享数据及数据共享服务过程进行监控，确保共享的数据未超出授权范围。
i) 建立共享数据格式规范，如提供机器可读的格式规范，确保高效获取共享数据。

6.5.2.2 增强要求

大数据服务提供者应：
a) 定期评估数据共享机制、服务组件和共享通道的安全性。
b) 配置专业数据共享机制或服务组件，明确数据共享最低安全防护基线要求。

6.5.3 数据发布安全

6.5.3.1 一般要求

大数据服务提供者应：
a) 建立数据资源公开发布的审核制度，严格审核数据发布业务符合相关法律法规要求。
b) 明确数据资源公开内容、适用范围及规范，发布者与使用者权利和义务。
c) 依法公开大数据服务相关数据资源公告、资格审查、成交信息、履约信息等数据发布信息。
d) 建立数据资源公开事件应急处理流程，包括保障处理流程快速有效的必要措施。
e) 建立数据资源公开数据库，通过大数据发布平台服务实现公开数据资源登记、用户注册等共享数据和共享组件的验证互认机制。
f) 指定专人负责数据发布信息的披露，并且对数据披露人员进行安全培训。
g) 定期审查公开发布的数据资源中是否含有非公开信息，并采取相关的措施，确保发布数据使用的合规性。

6.5.3.2 增强要求

大数据服务提供者应建立数据资源发布接口及发布数据格式规范，如提供机器可读的可扩展标记语言格式，确保用户能高效获取开放数据资源。

6.5.4 数据交换监控

6.5.4.1 一般要求

大数据服务提供者应：
a) 采用自动和人工审计相结合的方法或手段对高风险数据交换操作进行监控。
b) 记录数据交换操作事件，并制定数据交换风险行为识别和评估规则。
c) 部署必要的数据防泄漏实时监控技术手段，监控及报告个人信息、重要数据等的外发行为。
d) 使用数据处理平台对被监控的数据交换服务流量数据进行数据安全分析。

6.5.4.2 增强要求

大数据服务提供者应：

a) 记录数据交换服务接口调用事件信息，监控是否存在恶意数据获取、数据盗用等风险。

b) 具备对异常或高风险数据交换操作的自动化识别和实时预警能力。

6.6 数据销毁

6.6.1 介质使用管理

6.6.1.1 一般要求

大数据服务提供者应：

a) 建立大数据服务存储介质访问和使用安全策略和管理规范。

b) 从可信渠道购买或获得存储介质，采取有效的介质净化技术和规程对存储介质进行净化。

c) 对存储介质进行标记，明确介质存储的数据对象，并对介质访问和使用行为进行记录和审计。

d) 进行常规和随机检查，确保存储介质的使用遵守机构公布的关于介质的使用规范。

6.6.1.2 增强要求

大数据服务提供者应建立介质管理系统，确保存储介质的使用和传递过程得到跟踪。

6.6.2 数据销毁处置

6.6.2.1 一般要求

大数据服务提供者应：

a) 建立数据销毁策略和管理制度，明确销毁对象和流程。

b) 建立数据销毁审批机制，设置销毁相关监督角色，监督操作过程。

c) 依照数据分类分级建立相应的数据销毁机制，明确销毁方式和销毁要求。

d) 针对网络存储数据，建立硬销毁和软销毁的数据销毁方法和技术，如基于安全策略、基于分布式杂凑算法等网络数据分布式存储的销毁策略与机制。

e) 针对闪存、硬盘、磁带、光盘等存储数据，建立硬销毁和软销毁的数据销毁方法和技术。

f) 配置必要的数据销毁技术手段与管控措施，确保以不可逆方式销毁数据及其副本内容。

g) 按照国家相关法律和标准销毁个人信息、重要数据等敏感信息。

6.6.2.2 增强要求

大数据服务提供者应：

a) 建立数据销毁效果评估机制。

b) 对数据销毁效果进行认定。

c) 建立已共享或者已被其他用户使用的数据销毁管控措施。

6.6.3 介质销毁处置

6.6.3.1 一般要求

大数据服务提供者应：

a) 建立介质销毁处理策略、管理制度和机制，明确销毁对象和流程。

b) 依据介质存储内容的重要性明确磁介质、光介质和半导体介质销毁方法和机制。

c） 制定对存储介质进行销毁的监管措施，确保对销毁介质登记、审批、交接等介质销毁过程监控。

d） 按照国家相关法律和标准销毁存储介质，加强对介质销毁人员监管。

6.6.3.2 增强要求

大数据服务提供者应：

a） 使用国家权威机构认证的机构或设备对存储介质设备进行物理销毁。

b） 监控销毁过程，并对介质销毁效果进行认定。

c） 联系国家认定资质的销毁服务提供商执行存储介质销毁工作。

附 录 A
（资料性附录）
大数据服务模式、用户角色与业务目标

A.1 大数据服务模式

大数据系统主要由大数据平台和大数据应用两部分组成。大数据平台对不同大数据应用数据源进行聚合和融合、分布存储、并行处理和有效分析这些海量异构数据，并使用多种协议简化各数据源之间的数据接口、编程接口和数据提供者接口之间的映射，封装各种类型数据实体操作，以及可伸缩服务过程中的异常处理，使大数据使用者可以透明地访问或使用大数据系统中多源数据，以通用的、可互操作的、灵活的使用模式管理这些海量、异构、快速变化的数据资源。大数据应用为数据提供者和大数据使用者提供数据采集、数据传输、数据存储、数据处理（包括计算、分析、可视化等）、数据交换，直至数据销毁等覆盖数据生命周期相关数据活动的数据服务。

大数据服务存在数据自营模式、数据租售模式、数据仓库模式、数据平台模式、数据众包模式、数据外包模式等多种商业模式。不同商业模式下对大数据平台的基础设施资源选择、大数据计算和分析平台的资源调度与管理、支持数据生命周期管理的大数据应用数据服务组件部署等控制范围不同，从而大数据服务提供者在为大数据使用者提供大数据服务时所要求的安全能力也不同。

按照基础设施、数据平台与应用支撑的大数据服务层次结构，大数据平台服务可细分为通用计算资源/计算设施服务、大数据平台服务和大数据应用支撑服务（图 A.1 所示）。通用计算资源/设施服务一般包括计算硬件或虚拟机计算资源、网络通信资源和数据存储资源等组成，它为大数据平台与应用服务提供可伸缩的计算、通信和存储基础设施；大数据平台服务主要为海量、复杂、异构、动态变化的大数据计算与分析提供可扩展的各种数据处理和数据存储服务；大数据应用支撑服务主要指通过集成领域业务和数据模型、数据挖掘与分析套件、大数据可视化套件、集群自动化调度、面向应用的数据生命周期管理等大数据应用领域相关的服务组件和开发接口，用以简化大数据应用开发和部署。

从大数据平台提供者的大数据平台基础设施、数据计算和分析能力、应用支撑接口支持等业务模式和部署实施方式，大数据平台的系统服务分为四种类型：

——核心大数据服务：大数据平台提供者采用虚拟化技术、云计算技术或数据仓库技术，向大数据应用提供可扩展存储结构、分布式计算、内存计算等通用服务能力，以及数据存储管理和数据快速交互式分析处理能力。核心大数据服务提供关系、键值、文档、半结构化、文本、流数据等多种结构化数据和图像和音频/视频等类型的非结构化数据组织和访问服务，提供面向海量数据的分布式数据存储服务和可伸缩的数据并行处理和分析与可视化服务，并具有能与其他开源的通用基础设施、分布式计算资源进行数据存储和计算交互的能力，提供大数据应用开发接口和支持工具，以支持大数据应用提供者通过这些开发接口组合使用核心大数据服务功能来构建和部署他们所需的大数据应用服务。

——高性能大数据服务：面向批量数据分析、流式数据分析、海量数据联机事务处理等高可靠、高性能、高可用、可伸缩大数据存储和计算服务需求，在核心大数据服务基础上通过向下集成核心大数据服务所需的服务器、存储与网络设备、虚拟化软件等通用基础设施和计算资源，减少大数据服务基础设施部署和运维管理复杂度，简化大数据服务性能优化等问题。具备高性能大数据服务的大数据平台提供者一般都为大数据使用者提供集成的服务器、存储设备、操作系统、虚拟化管理软件、数据管理系统以及一些为数据查询、处理、分析用途而预先安装的数据服务组件，并提供应用编程接口、数据访问服务等应用开发环境和系统健康监测服务，大数据应

用开发者需使用这些个性化的编程接口与服务组件开发相应的大数据应用程序。

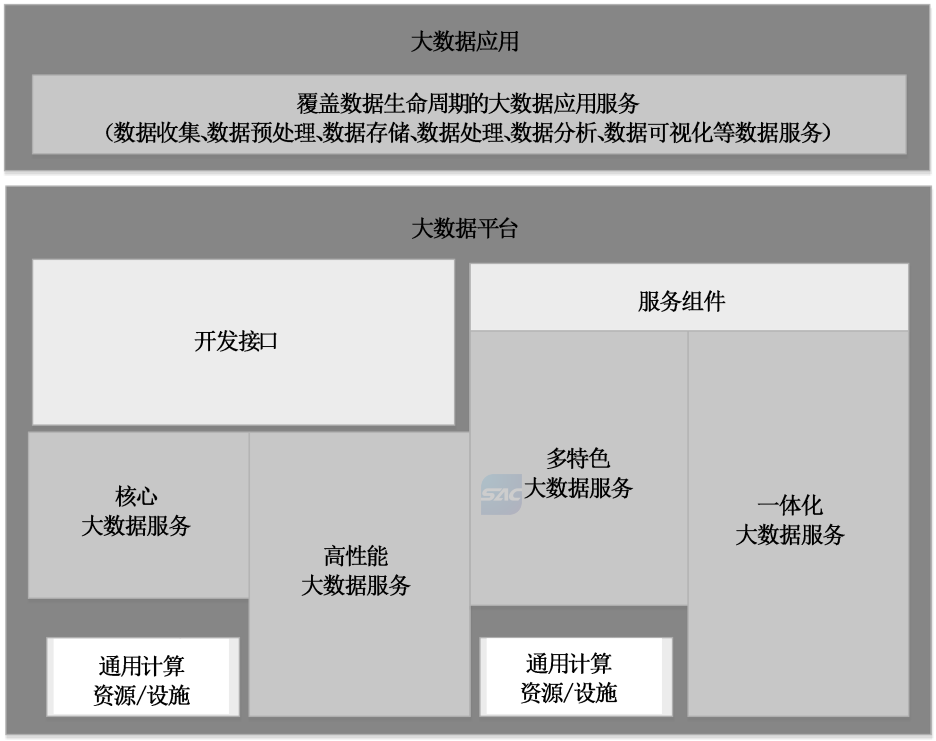

图 A.1　大数据服务类型和内容

——多特色大数据服务：面向金融、电信、能源、交通、电子政务等不同领域大数据服务共性需求，在核心大数据服务基础上通过向上集成领域相关的大数据分析与挖掘算法、业务数据模型、数据生命周期相关数据服务等面向应用的多种特征的数据业务服务，并提供包括应用编程接口、数据存储适配器等面向应用领域增强的特色大数据服务组件/构件。提供这种特色服务的大数据平台提供者一般在核心大数据服务中集成了大数据应用相关的数据服务基础功能，即提供支持大数据应用、具备领域特征的大数据建模、管理、处理和分析服务，使大数据应用提供者可借助这些特色大数据服务组件快速构建其大数据应用，启用大数据领域相关的特色服务。

——一体化大数据服务：大数据平台提供者在核心大数据服务基础上向下集成可扩展的大数据基础设施和通用计算资源，向上集成面向应用领域的数据采集、数据存储、数据分析、数据可视化等多特色大数据应用服务组件，一体化大数据服务平台提供者将核心大数据服务拓展为性能好、宜部署、大数据平台与大数据应用基础功能特色兼备的大数据存储和计算平台服务，为大数据使用者提供可扩展和完整的一站式大数据平台与应用支撑服务。

大数据服务提供者应通过大数据系统冗余和数据复制、备份等高可用解决方案，保证大数据服务水平协议的实现。例如基于云计算技术的大数据基础设施服务应参照 GB/T 31168—2014 及国家、行业或机构的有关信息安全标准规范落实相关的安全责任，以保证大数据服务的可扩展、可伸缩等服务可用性需求。此外，大数据服务提供者应该部署相关的服务安全管控组件，实时地监控大数据服务中数据主体、数据拥有者、大数据使用者及系统服务组件对大数据处理的各种属性，以保证实现大数据服务安全目标。

大数据服务提供者应依据其大数据应用服务目标和支撑大数据服务的大数据平台应担当的角色和责任，选择相应的大数据平台的系统服务类型，制定相应的安全策略和规范；识别机构大数据服务安全能力现状并分析与大数据安全目标的差距，挑选适合机构大数据服务数据业务和系统服务的安全控制

措施；在此基础上不断的改进大数据安全控制措施和制定大数据服务安全能力提升计划，可持续的保证大数据服务安全目标。

A.2 大数据用户角色

A.2.1 数据提供者

数据提供者将机构外部公共网络数据资源、外部合作企业私有数据资源、机构内部数据资源或大数据服务提供者系统运行过程中的各种日志、事件等系统行为数据资源进行抽象和建模，按照国家和行业数据安全标准与规范对这些数据源数据进行采集和整合后引入到大数据平台中，供大数据平台和大数据应用发现、访问、转换和分析这些数据资源。依据数据来源不同，数据提供者可进一步分为外部公共数据资源提供者、外部机构私有数据提供者、内部数据提供者、机器或系统数据提供者等数据提供角色。

A.2.2 大数据平台提供者

大数据平台提供者提供必要的网络、计算、存储等大数据服务所需的 IT 运行环境资源和必要的基础设施应用程序开发接口或服务组件，以支持大数据组织、存储、分析和基础设施部署和运维管理，响应大数据应用提供者提出的大数据服务请求。大数据平台提供者应通过大数据平台提供的身份标识与鉴别、授权与访问控制、密文处理与密钥管理、安全审计与数据溯源等安全功能保护数据处理的保密性和完整性、数据处理的真实性、数据处理过程中个人隐私保护等数据安全目标。鉴于大数据复杂性、多样性、快速变化等特点，为上层大数据应用提供分布式、可扩展的数据存储管理和分布式并行计算服务是大数据平台提供者的核心目标，这需要通过对存储资源和计算资源的高效管理和有效调度来实现。因此，大数据平台提供者可进一步分为大数据基础设施服务提供者、大数据存储管理服务提供者和大数据应用支撑服务提供者。

A.2.3 大数据应用提供者

大数据应用提供者将整合的大数据资源及其应用组件以软件服务方式部署到大数据平台上，并通过应用终端安全接入、数据分类分级、输入数据验证等安全策略配置和安全控制措施实施，给大数据使用者提供安全的数据组织、存储、分析和可视化服务。大数据应用提供者可分为数据、技术和服务三种产业链角色，其服务安全能力依赖于其商业模式所处的角色和义务，需要依据数据生命周期各阶段中所负责的数据活动进行定义，确保满足数据安全和隐私保护需求。大数据应用提供者应采用机器学习、数据挖掘等技术帮助大数据使用者开展诸如精准营销等各种分析与咨询服务等。

A.2.4 大数据服务协调者

大数据服务协调者规范和集成机构大数据服务所需的大数据平台和各类大数据应用数据业务活动，配置和管理大数据平台和大数据应用支撑安全功能组件，以构建一个可安全运行的大数据服务生态系统，确保大数据应用的各项数据服务能在大数据平台上安全高效地正确运行。大数据服务协调者负责为大数据服务组件分配对应的物理或虚拟节点，合理分配和调度大数据服务所需要的计算和存储资源，确保大数据服务运行效率达到所需要求，并通过资源的自动化按需分配，保证大数据服务可用性；通过不同的安全技术手段和安全措施，构筑大数据服务安全防护体系，实现覆盖硬件、软件和上层应用的安全保护；按照信息系统安全防护要求，从管理安全、网络安全、主机安全、应用安全、数据安全等方面来保证大数据服务平台的安全性；通过配置合理的大数据平台和数据容灾框架，提升大数据系统服务资源灾备和恢复能力。

A.2.5 大数据使用者

大数据使用者可以是一个真实的终端用户或机构角色，也可以是一个应用系统，它使用大数据平台提供者或大数据应用提供者提供的数据和服务。大数据应用和大数据平台提供给大数据使用者的数据应经过数据脱敏、数据合规性控制等安全控制措施，并通过授权和访问控制，以保证数据保密性、数据完整性和个人信息保护。大数据使用者的数据服务安全要求一般通过与大数据服务提供者的服务契约和服务水平协议体现，大数据服务水平协议中应规范性描述大数据应用服务各方面的安全属性，包括输入/输出等数据完整性和保密性属性，服务安全约束和响应时间等服务质量约束，以及在数据业务层面的诸多服务质量属性，如涉及的业务规则、数据依赖关系、时间/人员消耗可用性等。服务水平协议中还要规范描述大数据服务参与方相关的关系，如服务间依赖关系、数据服务和数据资源约束关系、数据服务和应用组件间关系、服务消息间关系等。

A.3 大数据服务业务安全目标

A.3.1 大数据应用、设备与外部服务组件安全管理

数据提供者和大数据使用者都是通过大数据应用终端、大数据服务组件和大数据服务接口与大数据系统进行交互。因此，大数据系统应安全地管理接入的大数据应用程序、大数据应用终端和外部潜在的大数据资源，提供诸如大数据应用程序安全注册、大数据应用安全元数据管理、大数据应用开发和部署策略等。大数据应用程序安全注册需登记和管理如物联网设备、移动终端等大数据应用终端设备、数字化产权保护下的各种数据资产，以及外部服务、应用程序和用户角色。大数据应用安全元数据管理应结构化存储和维护大数据服务安全相关的设备、用户、资产、服务组件等所有数据和主体安全要素，包括数据快速更新、数据结构变化，以及临时数据存储、数据有效性、大数据服务运行日志、溯源数据等系统运行安全统计数据，以支持应用数据生命周期、合规性控制等复杂应用的安全管理。大数据应用开发和部署实施策略涵盖符合机构信息系统环境建设的大数据应用部署和设施策略、大数据运行过程中的细粒度审计政策，以及不同大数据服务角色相关的行为规范等。大数据应用应该提供用户数据导入与导出，用户数据备份、用户行为数据保护等个性化数据安全管理功能。

A.3.2 大数据服务用户身份鉴别与访问控制

大数据应用提供者和大数据平台提供者应对大数据用户身份进行验证，并提供访问控制授权引擎对用户访问的数据资源进行控制，提供诸如基础设施层用户身份验证、应用程序层身份验证、终端用户层身份管理、服务提供商身份管理、粗粒度、细粒度、属性基等多种访问控制、多租户数据安全管理等。基础设施层用户身份验证应支持分布式计算技术、虚拟计算技术等计算方式的身份验证，例如基于硬件安全模块支持下的可信计算体系，从基础设施层提高大数据服务整体安全性。应用程序层用户身份验证应提供基于公钥基础设施等技术的身份认证服务平台，实现对应用层用户的证书、账户、授权、认证和审计的集中管理，实现整合大数据服务资源、应用数据共享和全面集中管控的目标。终端用户层身份管理应依据数据提供者和大数据使用者角色自动判断大数据应用中用户的身份信息，保证大数据系统中的用户标识和大数据应用用户参考标识与应用层授权信息之间的映射关系。服务提供商身份管理针对大数据系统中数据提供者、大数据服务协调者、大数据平台提供者、大数据应用提供者和大数据使用者，以多个服务身份使用大数据服务，使用安全性断言标记语言来定义数据资源提供者提供身份(和角色)，添加安全和隐私保证等要求，扩展传统的用户身份鉴别和授权机制。大数据可聚合多个数据提供者的数据资源，细粒度访问控制使得大数据服务提供者不只是分享数据集和数据服务，同时也分享数据授权策略。因此，大数据系统需要提供基于属性访问控制引擎，提供策略编辑点、策略决策点、策略执行点和策略访问点等面向数据对象的授权管理和访问控制功能。

A.3.3 大数据服务数据活动安全管理

围绕大数据的“数据—信息—知识—价值”数据价值链的数据生命周期活动，提供数据在传输、存储和使用过程中的加密功能和密钥管理功能，提供不同应用之间数据隔离与封装服务，确保用户数据保密性和可管理性；提供数据存储安全控制措施，包括不同数据副本或数据在不同空间的完整性检测措施，预防数据丢失，保证数据可访问性；部署必要的网络服务网关，确保数据的安全迁移、转换、交换和共享；提供聚合数据管理措施，确保多数据源安全整合；提供服务组件计算可信验证机制，确保应用服务组件的安全性；具备密文数据的透明计算能力；制定部署、迁移和保留策略、个人信息保护策略、去标识化和匿名化机制，确保数据生命周期中个人信息、重要数据等数据的安全管理，包括个人信息再标识风险管理等；提供大数据应用终端验证、数字版权管理、信任管理、数据披露、数据交易、数据治理等数据生命周期相关的是数据服务安全措施。

A.3.4 大数据服务基础设施安全管理

提供网络、计算、存储和环境资源，包括点对点传输、存储转发、大数据交换与通信框架操作和维护相关的安全措施和隐私保护功能，提供诸如威胁和脆弱性管理、安装与配置管理、系统监测和预警、运行日志和安全审计日志、网络边界控制和基础设施冗余和恢复等功能。威胁和脆弱性管理应识别大数据平台及大数据应用的脆弱性及相关威胁，对分布式拒绝服务攻击、密钥管理、加密协议，以及对脆弱性衍生的问题进行管理。安装与配置管理包括安全参数设置、安全组件部署、安全补丁管理、系统升级等，目的是保护大数据服务基础设施完整性。系统监测和预警需要通过部署大数据运行安全相关的组件和服务实现，它基于大数据基础设施运行数据实现大规模安全情报、复杂事件融合、安全分析、恶意软件监测和修复等安全功能。运行日志和安全审计日志是通过管理基础设施产生的海量、多样和高速变化的日志大数据，在线分析和统计抽样这些日志信息，为基础设施的安全运营和优化提供安全统计数据。网络边界控制主要为数据源和数据服务不可知的基础设施安全域间建立一条安全连接通道，共享服务网络体系结构，保证在开放环境下基础设施的网络通信安全。基础设施冗余和恢复通过系统复制有计划的维护大数据系统内部软件层次的冗余，以支持故障转移、系统恢复能力或减少大数据基础设施性能延迟，因为从大数据安全失败中恢复系统可能比传统集中式数据管理基础设施需要更加高级的基础设施安装、部署与配置等准备工作。

A.3.5 大数据系统应急响应管理

提供大数据服务基础设施、数据管理平台风险和责任相关的问题追责、安全合规、安全取证、安全事件管理、风险控制措施等。问题追责主要基于大数据平台和大数据应用之间的信息、流程和角色行为，通过追踪大数据系统的门户和检测点、向前和向后的溯源数据检查等方式实现。大数据系统安全和隐私的合规跨多个领域，涉及隐私、行业规范和本国的法律。安全取证可通过大数据安全分析服务组件取证，也可通过在大数据安全失败场景下取证。安全事件管理落实事件处理所需的各类支持资源，为用户处理、报告安全事件提供咨询和帮助。风险控制措施协调应急响应活动与事件处理活动，并与大数据服务相关外部机构（如供应链中的外部服务提供商等）提供事件应急处理机制。

参 考 文 献

[1] 《中华人民共和国网络安全法》中华人民共和国主席令第五十三号

[2] 《关于加强国家网络安全标准化工作的若干意见》中网办发文〔2016〕5 号

[3] 《国务院关于印发促进大数据发展行动纲要的通知》国发〔2015〕50 号

[4] 《国务院办公厅关于运用大数据加强对市场主体服务和监管的若干意见》国发〔2015〕51 号

[5] 《国务院关于大力推进信息化发展和切实保障信息安全的若干意见》国发〔2012〕23 号

[6] 《网络产品和服务安全审查办法(试行)》国家互联网信息办公室

[7] 《个人信息和重要数据出境安全评估办法(征求意见稿)》国家互联网信息办公室

[8] 《国家网络空间安全战略》国家互联网信息办公室

[9] NIST Special Publication 1500-2,NIST Big Data Interoperability Framework:Volume 2,Big Data Taxonomies, September 16,2015

[10] NIST Special Publication 1500-4,NIST Big Data Interoperability Framework: Volume 4, Security and Privacy,September 16,2015

[11] NIST Special Publication 1500-6,NIST Big Data Interoperability Framework: Volume 6, Reference Architecture,September 16,2015