

# 突破大数据发展的瓶颈

韩涵

13811249349 , hanhan@ritt.cn

中国信息通信研究院

2017年6月7日

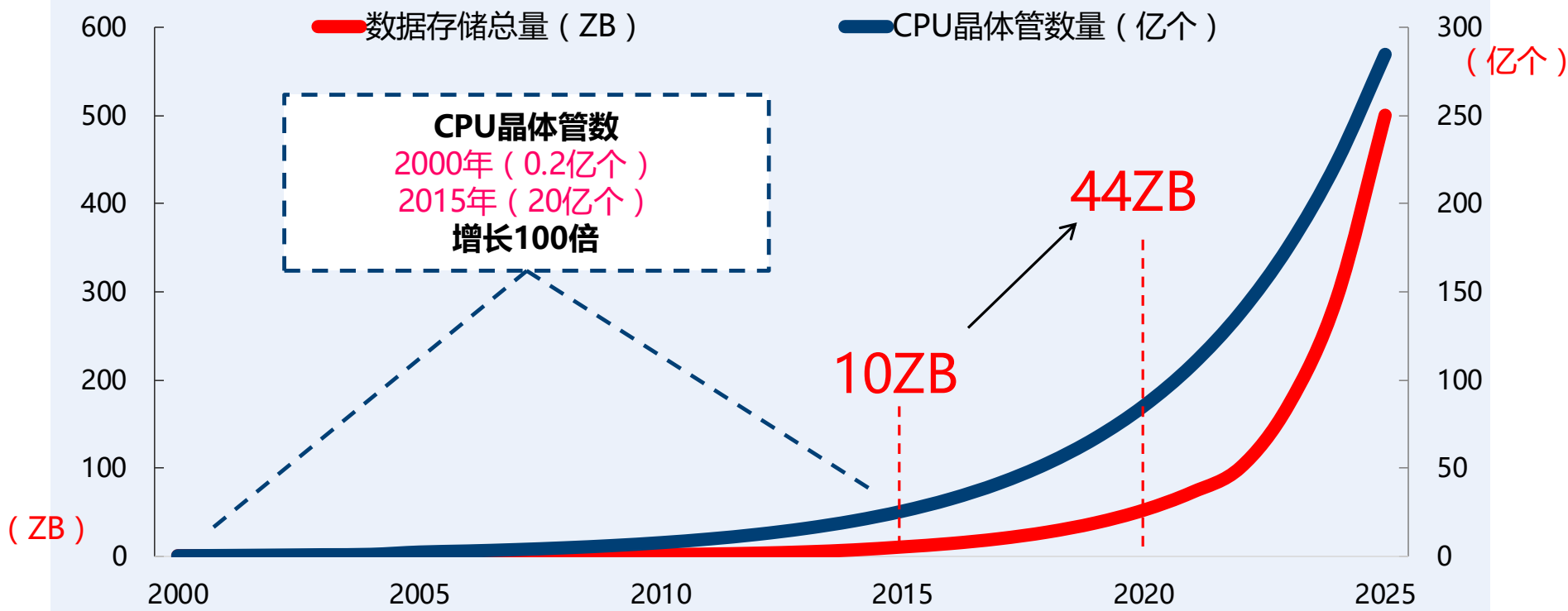


- 一、国内外大数据发展政策**
- 二、发展大数据的基本理论问题**
- 三、发展大数据的战略选择**

# 万物互联时代 数据爆炸式增长

全球互联网步入泛在普及、深度融合、变革创新、引领转型的新阶段

- 数据存储量指数级增长。2015年，10ZB，2020年，达44ZB，是15年的4倍多
- 计算机处理能力按摩尔定律增长。2000-2015年CPU晶体管数量增加100倍
- 存储设备价格大幅下降。2000-2015年硬盘价格降低为1/200

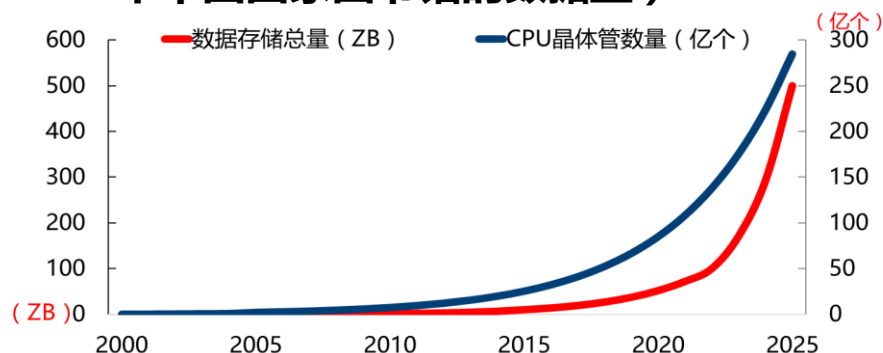


# 大数据的特征

- 人类社会已经进入大数据时代，全球数据量每18个月翻一番
- “大数据”定义目前尚无统一定论，业界多数认为大数据是**4V特征的数据集合**

## Volume数据量大

2020年全球数据总量将达44ZB  
(1ZB=1万G比特字节，相当于776万个中国国家图书馆的数据量)



## Variety类型复杂

2020年我国非结构化数据占比达到90%



## Velocity速度快

实时快速数据处理、实时结果导向



## Value价值

运用非常规工具完成数据处理，产生巨大价值



# 各主要国家积极布局大数据

——发达国家均高度重视大数据在促进经济社会创新变革、提升国家竞争力中的重要作用，出台国家战略，落实配套措施，系统推动大数据发展



美国认为大数据事关国家竞争力，多措并举推动发展

开放数据、资助研发、带头应用、隐私保护



欧盟：2014年  
出台数据驱动经济战略



澳大利亚：2013年  
发布了公共服务大数据战略



新加坡：建设数据地平线平台（RAHS），数据治国

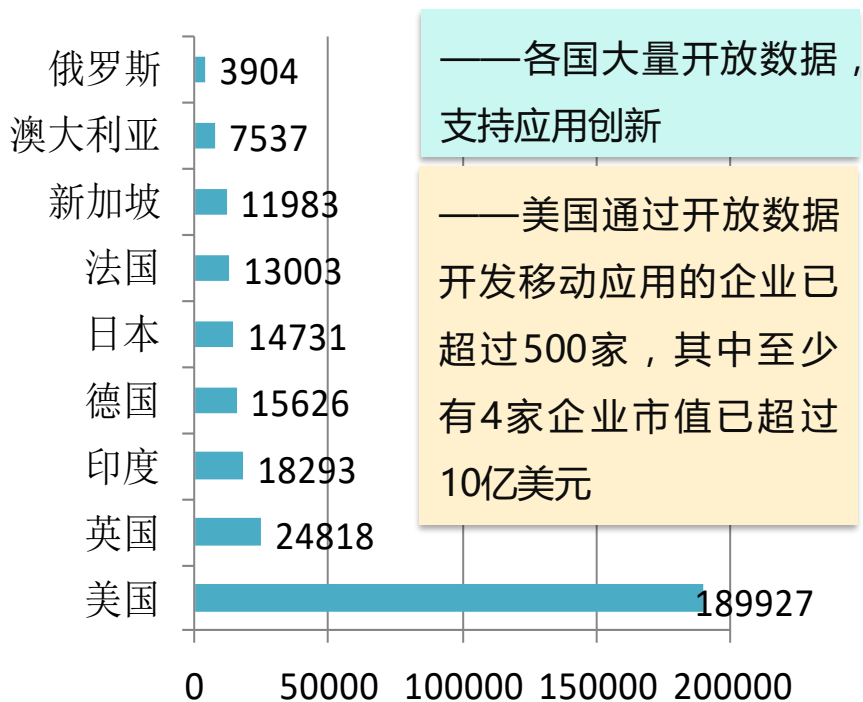


日本发布  
《创建最尖端IT国家宣言》

# 发达国家推进大数据发展的主要方向（1）

——发达国家从推进数据开放、加强技术研发、构建产业生态和完善法律法规四方面促进大数据发展

**1、数据开放：**实施数据开放的国家达97个，支持大批创新应用和创业



各国数据开放网站开放数据集量（个）（截至2015.10.29）

**2、技术研发：**加大投入支持基础、核心、前沿、公共应用研发



# 发达国家推进大数据发展的主要方向（2）

## 3、产业生态：强化公私合作和跨部门协作，促进形成大数据生态体系

美



研发支持中注重多部门协同，强调与大学、研究机构和企业协同

## 4、法律法规：立法探索，以应对个人隐私保护、知识产权保护等新问题

——发布白皮书，对保护个人隐私和公平权利提出建议

欧



与行业协会签署PPP协议，制定政策、运作项目，引导产业投资、研发创新

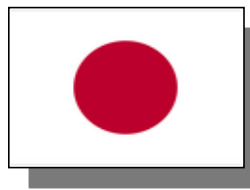
——探索解决个人数据保护、文本挖掘著作权保护、数据所有权建立等问题

澳



通过公私合作促进公共服务大数据发展，扶持商业创新

日



修订《个人信息保护法》，允许企业向第三方出售充分匿名化数据

# 习近平：建设全国一体化的国家大数据中心



以推行电子政务、建设新型智慧城市等为抓手，以数据集中和共享为途径，建设全国一体化的国家大数据中心，推进技术融合、业务融合、数据融合，实现跨层级、跨地域、跨系统、跨部门、跨业务的协同管理和服务。

——2016年10月9日，习近平主持第三十六次中共中央政治局集体学习时发表讲话

要着力推动互联网与实体经济深度融合发展，以信息流带动技术流、资金流、人才流、物资流，促进资源配置优化，促进全要素生产率提升，为推动创新发展、转变经济发展方式、调整经济结构发挥积极作用。

——2016年4月19日，习近平主持召开网络安全和信息化工作座谈会并发表讲话



# 李克强：大数据是“钻石矿”



要进一步推进“互联网+”行动，广泛运用物联网、大数据、云计算等新一代信息技术，促进不同领域融合发展，催生更多的新产业、新业态、新模式，推出更加符合市场需要的新产品和服务。

——2016年6月27日，李克强总理在第十届夏季达沃斯论坛上的演讲

大数据等新一代互联网技术深刻改变了世界，也让各国站在科技革命的同一起跑线上。中国曾屡次与世界科技革命失之交臂，今天要把握这一历史机遇，抢占先机，赢得未来。

——2016年5月25日，李克强总理在贵阳出席中国大数据产业峰会上的演讲

深入实施《中国制造2025》，加快大数据、云计算、物联网应用，以新技术新业态新模式，推动传统产业生产、管理和营销模式变革。”。

——2017年3月5日，李克强总理在十二届全国人大五次会议上作政府工作报告，这是“大数据”第四次写入政府工作报告

# 我国实施国家大数据战略

## 《中华人民共和国国民经济和社会发展第十三个五年规划纲要》第二十七章：实施国家大数据战略

### 第一节 加快政府数据开放共享

- 全面推进重点领域大数据高效采集、有效整合。
- 加快建设国家政府数据统一开放平台。
- 制定政府数据共享开放目录，依法推进数据资源向社会开放。

### 第二节 促进大数据产业健康发展

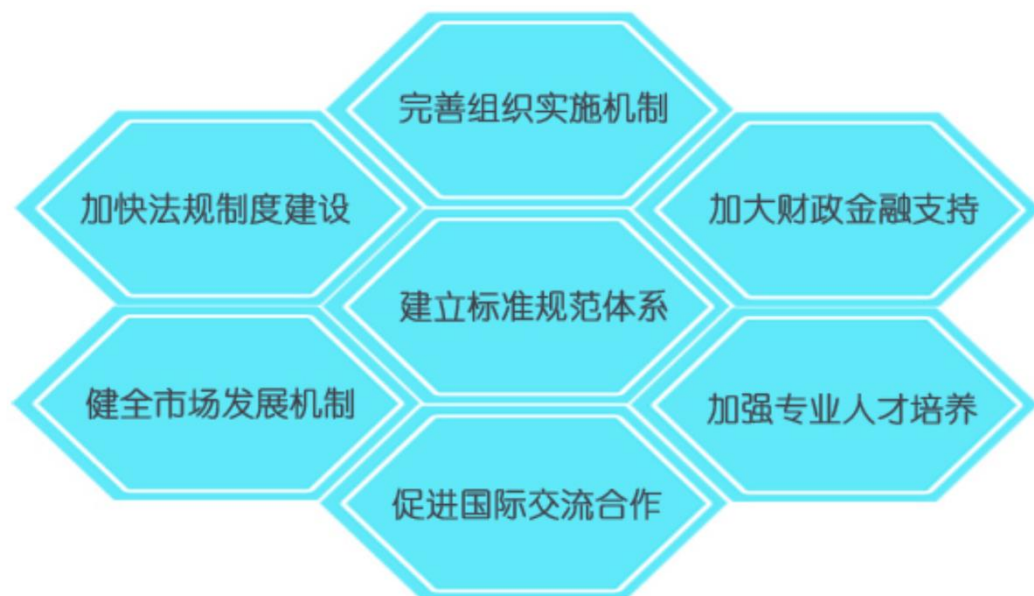
- 深化大数据在各行业的创新应用，探索与传统产业协同发展新业态新模式，加快完善大数据产业链。
- 加快海量数据采集、存储、清洗、分析发掘、可视化、安全与隐私保护等领域关键技术攻关。
- 促进大数据软硬件产品发展。
- 完善大数据产业公共服务支撑体系和生态体系，加强标准体系和质量技术基础建设。



# 国家顶层设计：《促进大数据发展行动纲要》

2015年8月31日，国务院正式印发《促进大数据发展行动纲要》（国发〔2015〕50号），从总体目标、主要任务、政策机制等方面全面部署了我国大数据的发展。为国务院各部委、地方政府制订大数据行业发展规划提供了指导思想和重要依据。

——中国信息通信研究院是《行动纲要》出台的核心支撑单位



# 《促进大数据发展行动纲要》 内容介绍

## 《促进大数据发展行动纲要》

一个目标

全面推动我国大数据发展和应用，加快建设数据强国

三大内容

加快政府数据开放  
推动资源整合  
提升治理能力

推动产业创新发展  
培育新业态  
助力经济转型

健全安全保障体系  
强化安全支撑  
提高管理水平

十项工程

政府  
数据  
资源  
共享  
开放  
工程

国家  
大数  
据资  
源统  
筹发  
展工  
程

政府  
治理  
大数  
据工  
程

公共  
服务  
大数  
据工  
程

工业  
和新  
兴产  
业大  
数据  
工程

现代  
农业  
大数  
据工  
程

万众  
创新  
大数  
据工  
程

关键  
技术  
及产  
品研  
发与  
产业  
化工  
程

大数  
据产  
业支  
撑能  
力提  
升工  
程

网络  
和大  
数据  
安全  
保障  
工程

七条措施

完善组织实施机制

加快法规制度建设

健全市场发展机制

建立标准规范体系

加大财政金融支持

加强专业人才培养

促进国际交流合作

# 工信部：大数据产业发展规划（2016-2020）

为全面部署“十三五”时期大数据产业发展工作，加快建设数据强国，为实现制造强国和网络强国提供强大的产业支撑，工业和信息化部于2016年12月公布了《大数据产业发展规划（2016-2020年）》。规划提出了我国大数据产业发展总体目标及五个分目标为：

**到2020年，技术先进、应用繁荣、保障有力的大数据产业体系基本形成。大数据相关产品和服务业务收入突破1万亿元，年均复合增长率保持30%左右，加快建设数据强国，为实现制造强国和网络强国提供强大的产业支撑。**

**——中国信息通信研究院是《规划》编制的牵头支撑单位**

**技术产品先进可控**

**应用能力显著增强**

**生态体系繁荣发展**

**支撑能力不断增强**

**数据安全保障有力**

# 大数据产业发展规划：主要内容



# 主要任务1——强化大数据技术产品研发

以应用为导向,突破大数据共性关键技术,推动产品和解决方案研发及产业化,不断创新技术服务模式,形成技术先进、生态完备的技术产品体系。

加快大数据关键  
技术研发



培育大数据产  
品体系



创新大数据技  
术服务模式

## 专栏 1：大数据关键技术及产品研发 与产业化工程

突破技术

打造产品

树立品牌

## 专栏 2：大数据服务能力提升工程

**培育数据即服务模式。**发展数据资源服务等模式

**支持第三方大数据服务。**培育一批开展数据服务的新业态。



## 主要任务2——深化工业大数据创新应用

加强工业大数据基础设施建设规划与布局,推动大数据在产品全生命周期和全产业链的应用,形成数据驱动的工业发展新模式,支撑中国制造2025战略。

加快工业大数据基础设施建设

推进工业大数据全流程应用

培育数据驱动的制造业新模式

### 专栏 3：工业大数据创新发展工程

加强工业大数据关键技术研发及应用

建设工业大数据公共服务平台

重点领域大数据平台建设及应用示范

探索工业大数据创新模式



# 主要任务3——促进行业大数据应用发展

加强大数据在重点行业应用，促进跨行业大数据融合创新，在政府治理和民生服务中提升大数据运用能力，推动大数据与各行业领域的融合发展。

重点行业大数据应用

跨行业大数据融合创新

社会治理和公共服务  
大数据应用

## 专栏 4：跨行业大数据应用推进工程

跨行业大数据试  
点示范

跨行业大数据推  
进组织

大数据融合应用试  
验床

# 主要任务4——加快大数据产业主体培育

引导区域大数据发展布局,促进基于大数据的创新创业,培育一批大数据龙头企业和创新型中小企业,形成多层次、梯队化的创新主体和合理的产业布局,繁荣大数据生态。

大数据助推创新创业

构建企业协同发展格局

优化产业区域布局

## 专栏 5：大数据产业集聚区创建工程

围绕大数据产业集聚区创建系列工作,明确工作机制,进行总体设计和具体指导。建立科学、合理、客观、可操作的大数据产业集聚区创建指标体系,明确创建标准。

# 主要任务5——推进大数据标准体系建设

加强大数据标准化顶层设计,出台急用先行标准,逐步完善标准体系,发挥标准化对产业发展的重要支撑作用。

加快重点标准研制与推广

积极参与大数据国际标准化工作

## 专栏 6：大数据重点标准研制及应用示范工程

国家标准研制

建立验证和检测平台

开展标准应用示范

# 主要任务6——完善大数据产业支撑体系

统筹布局大数据基础设施,建设大数据产业发展创新服务平台,建立大数据统计及发展评估体系,创造良好的产业发展环境。

合理布局大数据基础设施建设

构建大数据发展公共服务平台

建立大数据发展统计评估体系

## 专栏 7：大数据公共服务体系建设工程

支持建立大数据产业公共服务平台

支持第三方机构建立测试认证平台

支持建立大数据开源社区

# 主要任务7——提升大数据安全保障能力

应对大数据面临的网络信息安全新形势,加强大数据安全技术产品研发,利用大数据完善安全管理机制,构建强有力的大大数据安全保障体系。

加强大数据安全技术产品研发

提升大数据对网络信息安全的支撑能力

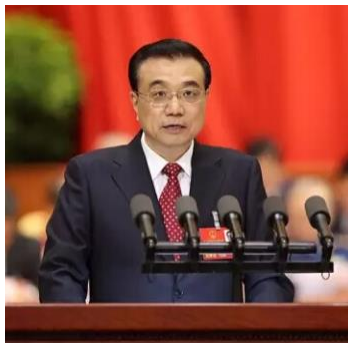
## 专栏 8 : 大数据安全保障工程

开展大数据安全产品研发  
与应用示范



支持建设一批大数据  
安全攻防仿真实验室

# 国家大数据综合试验区：大数据发展需要示范引领



“坚持创新驱动发展，加快大数据部署，深化大数据应用，已成为稳增长、促改革、调结构、惠民生和推动政府治理能力现代化的**内在需要和必然选择**。”

“开展区域试点，推进**贵州等大数据综合试验区**建设，促进区域性大数据基础设施的整合和数据资源的汇聚应用。”

——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)



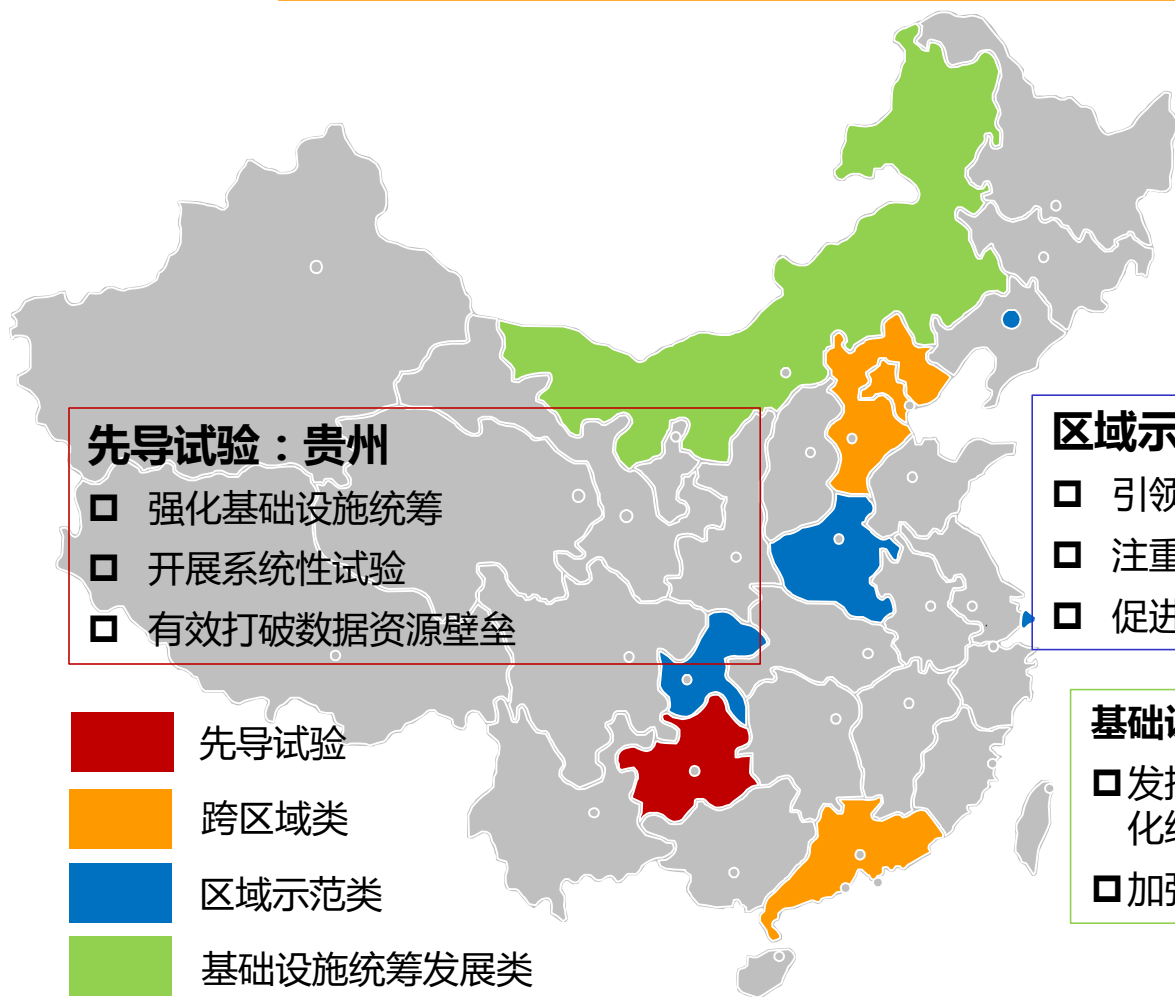
## 试验目标

- ❑ 通过3~5年时间探索，推动制度、管理、服务和协同创新，有效打破数据资源壁垒，形成若干辐射带动效应强、示范引领作用显著的区域性试验基地
- ❑ 打造一批大数据先进产品，培育一批大数据骨干企业，建设一批大数据众创空间，培养一批大数据产业人才，有效推动相关制度创新和技术创新，发掘数据资源价值，提升政府治理能力，推动经济转型升级

# 全国多片区布局综合试验区

## 批复 历程

- 2016年2月25日，三部委批复**贵州**建设国家大数据(贵州)综合试验区。
- 2016年10月8日，三部委批准**7个区域**建设国家大数据综合试验区。



# 全国各地方纷纷制订各自大数据发展政策规划

据中国信息通信研究院统计，截止到2017年1月底，共有16个省、21个市发布了大数据发展规划

省市	文件名称	省市	文件名称	省市	文件名称
北京	北京市大数据和云计算发展行动计划	广州	关于促进大数据发展的实施意见	沈阳	沈阳市促进大数据发展三年行动计划
上海	上海市大数据发展实施意见	深圳	深圳促进大数据发展行动计划	哈尔滨	促进大数据发展若干政策（试行）
江苏	江苏省大数据发展行动计划	青岛	关于促进大数据发展的指导意见	合肥	合肥市大数据发展行动纲要
浙江	浙江省促进大数据发展实施计划	宁波	关于推进大数据发展的实施意见	淮南	大数据产业发展三年行动计划
广东	广东省促进大数据发展行动计划	南京	南京促进大数据发展三年行动计划	东莞	东莞市大数据发展实施方案
山东	关于促进大数据发展的意见	武汉	武汉大数据产业发展行动计划	兰州	促进大数据发展的实施意见
安徽	“十三五”软件和大数据产业规划	郑州	促进大数据发展行动计划	赣州	加快大数据发展实施意见
广西	促进大数据发展的行动方案	南宁	南宁市大数据建设发展规划	盐城	推进大数据产业发展的实施意见



# 全国各地确立符合发展情况的大数据发展定位

## 积极面向全国需求推进大数据应用与产业创新

	面向国家	面向地方
产业中心/高地	北京、重庆、 <b>广东</b> 、浙江、湖北、新疆、深圳、武汉、沈阳、宁波、天津滨海新区、贵安新区	青岛、合肥、淮南、台州
应用示范中心	北京、贵州、 <b>广东</b> 、浙江、湖北、深圳、青岛、宁波、淮南、贵安新区	青海、台州
资源聚集中心	贵州、湖北、青海、武汉、天津滨海新区	沈阳、兰州、淮南
创业创新中心	北京、 <b>广东</b> 、青岛、宁波、贵安新区	淮南
人才发展中心	湖北	

- 在发布大数据规划的37个省市中，明确定位的有20个，占统计总量的54%
- 12个省市提出建设面向全国的大数据产业中心或高地，产业发展受关注度最高
- 东部、中部、西部不同区域的省市均把大数据作为新的经济与产业增长点

# 全国各地区域发展特色与优势

## 五大区域构成我国大数据产业总视图

### 中西部地区

- 借助环境、能源、价格优势，实现集群效应，打造经济发展新动能
- 建设数据中心、网络基础设施与大数据产业园区

### 珠三角地区

- 以广州、深圳为引领，打造全国大数据应用先导区，建设全国的大数据产业聚集区，探索大数据的新路径、新经验

八大综合试验区，引领大数据发展新思维

内蒙古：大数据基础设施统筹发展综试区

京津冀：跨区域综试区

沈阳

河南

上海

重庆

贵州：首个综试区

珠三角：跨区域综试区

### 东北地区

- 以大数据发展为主体，以传统产业转型升级和智慧城市建设为两翼，完善大数据产业链，振兴东北老工业基地

### 环渤海地区

- 发挥北京的人才与科技优势，天津、河北、山东的制造业与资源优势，建设国家大数据产业创新中心、全球大数据产业创新高地

### 长三角地区

- 依托上海、江苏、浙江、安徽等地在金融、ICT、互联网、制造业等领域优势，实现大数据与智慧城市、云计算协同发展

# 广东省发布促进大数据发展行动计划

**总目标：用5年左右时间，打造全国数据应用先导区和大数据创业创新集聚区，抢占数据产业发展高地，建成具有国际竞争力的国家大数据综合试验区。**

**大数据基础设施建设、资源整合和政府数据开放共享取得显著成效。**

建成全省统一的电子政务数据中心，以及10个左右地市级政务数据分中心。

**大数据创新应用深入经济社会各领域。**

运用大数据推动形成精准治理、多方协作的社会治理新模式

**大数据产业成为重要的经济增长极。**

基本形成高端智能、新兴繁荣的大数据产业发展新生态和大众创业、万众创新的创新驱动新格局

**一、国内外大数据发展政策**

**二、发展大数据的基本理论问题**

**三、发展大数据的战略选择**

# 基本理论：数据的资产属性

权

利

责

数据的权利属于谁？

——数据具有资产的属性，国内外相关政策，均赋予数据这一属性

## 政府层面来看——政策法规

美



美国国家档案与文件管理局明确规定要把“数据”当作“**组织的信息资产**”来进行管理，**与管理资产的业务具有同等重要的地位**

英



英国BSI定义：**信息是一种资产**，像其他重要的业务资产一样，**对组织具有价值**

中



《关于加强信息资源开发利用的若干意见》（中办〔2004〕34号）提出“**信息资源作为生产要素**、无形资产和社会财富，与能源、材料、资源同等重要”

## 市场层面来看——数据商品化现象

### 数据交易所



### 互联网模式

例如：互联网广告营销中隐含数据价值

### 出售个人数据

2014年4月，荷兰学生肖恩·巴克斯建立一个拍卖网站来出售自己的个人数据，包括个人住址、医疗记录、社交网络信息等，最终他以350欧元的价格出售了**个人数据**。

# 基本理论：数据的权力归属问题

权

利

责

数据的权利属于谁？

从世界范围来看，**公共信息资源**归全民所有，开放是一种义务

- **美国**：发布《联邦政府信息资源管理政策》，**政府数据**是全民**有价资产**
- **英国**：政府开放数据的版权归皇家（虚拟概念）所有，**广大使用者**均可获得“**开放政府授权**”

从财产权、人格权来看，用户享有**个人数据**的所有权

处理个人信息所有权的两大出发点包括：“财产权”、“人格权”，根据**个人数据保护法**，**数据主体对于个人数据具有完整的控制权利**。

从投入成本来看，**企业对于去隐私的数据集**享有所有权

**一是**企业对于数据的记录投入了巨大的技术、网络、人力、管理成本；**二是**为企业利用数据创造财富提供保障。**三是**有利于规范数据交易市场，遏制数据的非法黑市交易，让数据在有序可控的规则之下充分流动。

# 基本理论：数据的权力归属问题

权

利

责

数据的权利属于谁？

政府

事业单位

国企

其他企业  
和个人

关系国家安全的  
数据

关系商业秘密的  
数据

关系个人隐私的  
数据

不含以上内容的  
数据

**我国确认数据权利归属面临的困难**

**1、——数据归属单位不清。**和国外不同，我国大量公共信息资源实际掌握在公共部门手中，公共部门为数据的采集、软硬件系统、运维付出了一定的成本，甚至为政府部门维护数据库。

按照网信办的内部指导意见，我国公共信息资源责任的主体包括4类：1）负有公共事务管理职能和公共服务职能的政府部门、2）市政公用企事业单位、3）公共服务事业单位、4）经政府部门授权实施公共服务的企业单位。

**2、——缺乏完善的个人信息保护法律体系。**  
个人数据、企业数据、公共数据的边界不清晰

# 基本理论：数据的权力归属问题

权

利

责

数据的权利属于谁？

**公共数据资源**  
政府+公共事业数据

**非公共数据资源**  
其他企业和个人数据

关系国家安全的  
数据

关系商业秘密的  
数据

关系个人隐私的  
数据

不含以上内容的  
数据

应以政府主导的手段**有**  
**限制的**内部**共享**，但不  
适合对外开  
放

不应流动

应以**政府主**  
**导**的手段开  
放共享

应以**市场主**  
**导**的手段流  
通交易

因此，可以将**数据权利的归属**  
**大致归为两类**：

- 1) **公共数据资源**：政府或公共事业单位负责，**属于全民所有**
- 2) **非公共数据资源**：企业或个人**私有**

**纳入公共数据资源的两条原则**：

- 1) 用法律或行政手段及公共财政支持采集的数据或通过政府特许经营的权利采集的数据
- 2) 数据来源的业务部门属于公共事务领域

**公共数据资源开放的原则**：  
不公开对民众公共利益产生不良影响



# 基本理论：数据的利益分配问题



数据带来的利益如何分

以数据作为**资产的角度**，评估价值的方法主要包括：**成本法、市场法、收益法**

- **成本法**：以资产建立时软硬件、人员等各项投入计算，补充历史和现实的差额以及长期运行费用
- **市场法**：以当前市场相同或类似的已交易资产作为参照物
- **收益法**：以预期收益折现评估资产

**美、欧**两大体系政府数据收益法则**不同**，我国以**美国模式**为主

- **美国模式**（**成本法模型** cost model）：美国立法规定政府对其拥有的信息资源收取的费用不得超过提供信息的成本价，只能包括搜索信息、

投资价值	经济价值
欧盟为典型的成本补偿模式（cost-recovery model）：政府信息生成机构直接从用户与其他机构处收回成本，每年957亿欧元	美国为典型的收益模式（benefit model）：政府信息生成机构需要为获取数据支付一定的费用。政府也参与商业性开发。每年1904亿欧元

我国的传统做法更像美国模式

# 基本理论：数据的利益分配问题

权

利

责

数据带来的利益如何分配？

公共数据资源

非公共数据资源

关系国家安全的数据

关系商业秘密的数据

关系个人隐私的数据

不含以上内容的数据

以成本法评估内部共享的成本，给与财政补贴

仅能内部使用，不能对外获取收益

以成本法评估给与财政补贴

以市场法、收益法获取交易收益

数据利益的问题可以分为两类：

1) **公共数据资源**，适用于**成本法**，由国家财政给与**补贴**。包括：建设成本、运维成本、历史与现实差价、为开放共享升级系统的成本等；**寻求市场化价值也应仅收取成本的部分**。

2) **非公共数据资源**：适用于**市场法**（交易）、**收益法**（权益类交易），**由市场的手段获取收益**。将在后面数据交易部分详细阐述。

# 基本理论：数据的利益分配问题

权

利

责

数据带来的利益如何分配？

公共数据资源

非公共数据资源

关系国家安全的数据

关系商业秘密的数据

关系个人隐私的数据

不含以上内容的数据

以成本法评估内部共享的成本，给与财政补贴

仅能内部使用，不能对外获取收益

以成本法评估给与财政补贴

以市场法、收益法获取交易收益

**终极原则**

——**鼓励基础数据开放**。公共数据资源的开放，应以基础性、原始性数据为主要原则，公共事业单位开放数据资源为基础性服务，即义务——**鼓励免费开放服务**。采用免费或低费方式，数据资源的再开发利用应鼓励商业机构进行；公共事业单位提供数据资源的深入开发为增值性服务，这部分可以收取商业利益，但不应鼓励。不公开的公共信息资源不建议收取增值费用。

# 基本理论：数据的责任问题

权

利

责

序号	法律法规	颁布机构	施行时间
1	《刑法》（2011）修正	全国人大常委会	2011年5月1日
2	《侵权责任法》	全国人大常委会	2010年7月1日
3	《反不正当竞争法》	全国人大常委会	1993年12月1日
4	《消费者权益保护法》（2014）修订	全国人大常委会	2014年3月15日
5	《关于加强网络信息保护的决定》	全国人大常委会	2012年12月28日
6	《关于审理利用信息网络侵害人身权益民事纠纷案件适用法律若干问题的规定》	最高人民法院	2014年10月10日
7	《征信业管理条例》	国务院	2013年3月15日
8	《征信机构管理办法》	中国人民银行	2013年12月20日
9	《侵害消费者权益处罚办法》	国家工商行政管理总局	2015年3月15日
10	《关于禁止侵犯商业秘密行为的若干规定》	国家工商行政管理总局	1998年12月3日
11	《电信和互联网用户个人信息保护规定》	国家工业和信息化部	2013年9月1日
12	《规范互联网信息服务市场秩序若干规定》	国家工业和信息化部	2012年3月15日
13	《关于加强网络团购经营活动管理的意见》	国家工商行政管理总局	2012年3月12日
14	《人口健康信息管理办法（试行）》	国家卫生计生委	2014年5月5日
15	《信息安全技术公共及商用服务信息系统个人信息保护指南》	国家工业和信息化部	2013年2月1日

# 基本理论：数据的责任问题

权

利

责

对数据的责任包括哪些？

## 一、首要责任是个人信息保护

### ③初始为模糊期，最终形成白名单或黑名单

- 在各国数据开放的初期，存在着“模糊期”和“试探期”，这一期间，数据评估主要依靠各部门在总体原则下进行裁量，判断数据开放的范围和内容。
- 经过一段时间的探索，目前美国、英国、新加坡等国均要求数据开放部门对数据进行评估分类，并对开放目录或负面清单负责。一般初期形成“白名单”，最终形成“黑名单”

### ④国内外数据开放网站均设立了“免责条款”



### ⑤推动个人信息保护最终要依靠立法，地方法律（政府令）可先行

# 基本理论：数据的责任问题



对数据的责任包括哪些？

## 二、其次是数据全生命周期的管理责任

2013年5月9日，美国奥巴马总统发布行政令《将开放、机器可读作为政府信息的新的默认状态》，提出**政府信息应该作为一种资产进行全生命周期管理**，促进其互操作性和开放性。

数据全生命周期包括：



数据全生命周期资产管理的内容：

- 1) 保护个人信息
- 2) 做好规范记录（行为的原因、行为的内容、行为的结果）
- 3) 经常比对结果

**一、国内外大数据发展政策**

**二、发展大数据的基本理论问题**

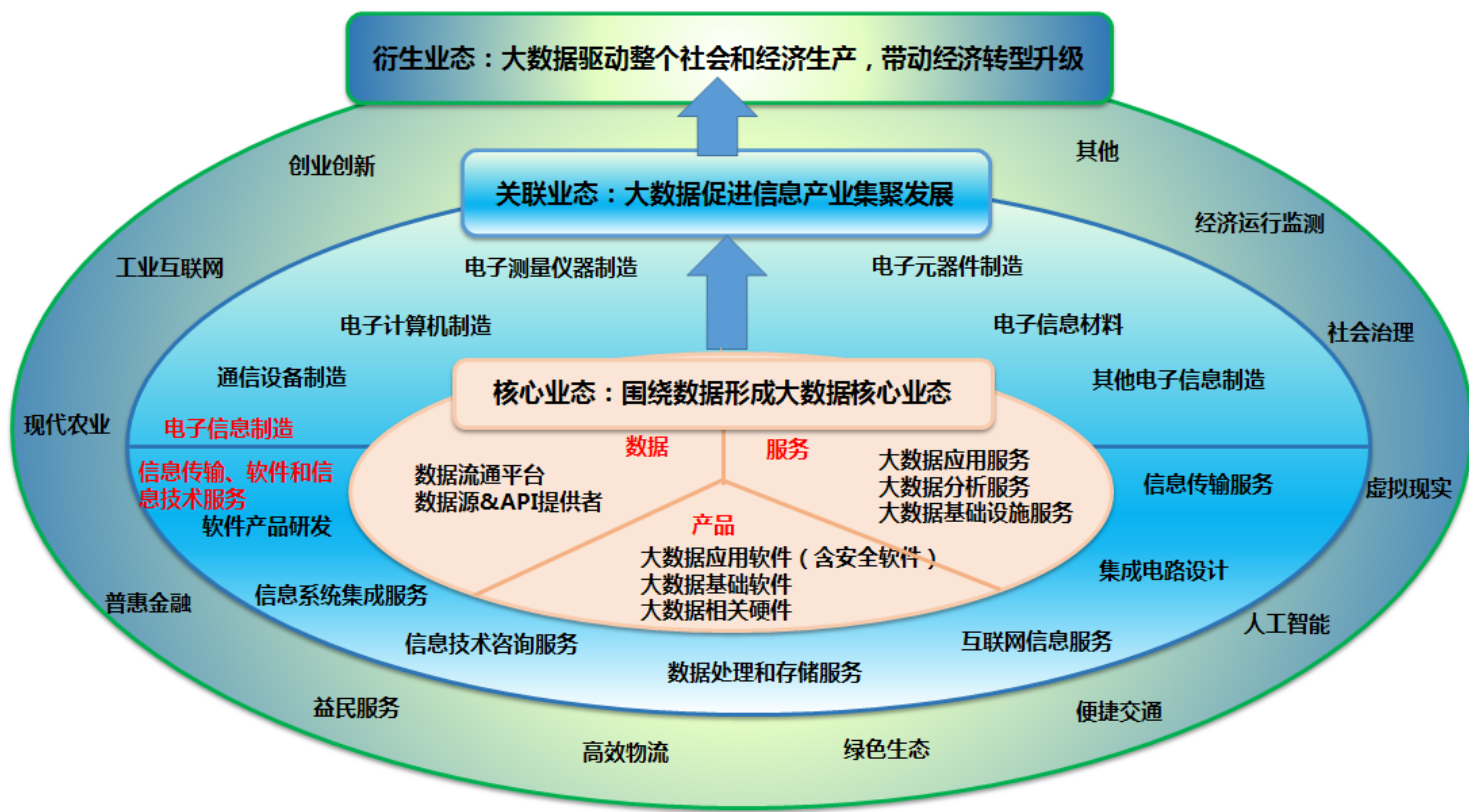
**三、发展大数据的战略选择**



# 《大数据产业发展规划》对大数据产业界定

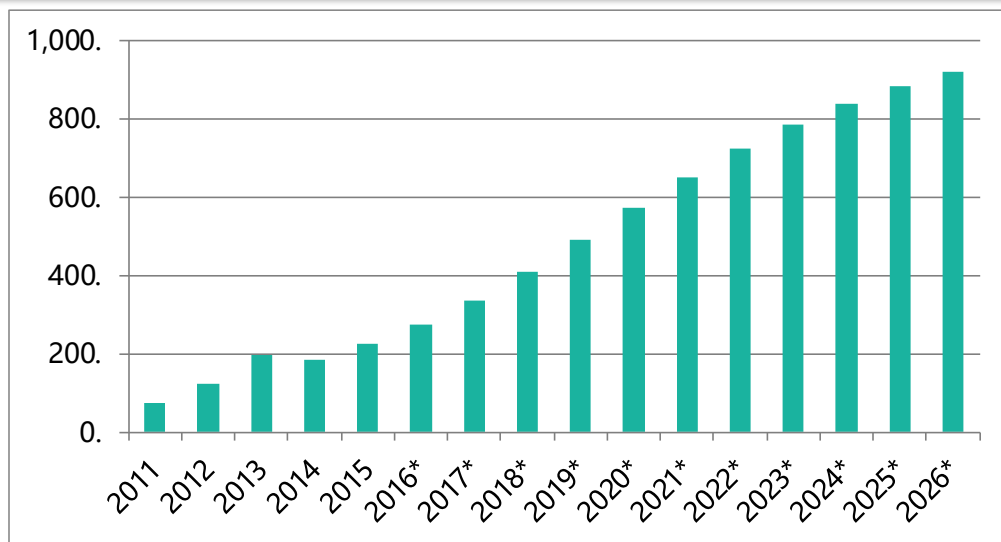
大数据产业由大数据核心业态、大数据关联业态和大数据衍生业态构成

- 核心业态，包含数据资源提供、大数据管理和分析软硬件及服务等环节
- 关联业态，对大数据提供基础支撑的电子信息产业，包括电子信息制造，信息传输、软件和信息技术服务等
- 衍生业态，通过大数据技术与传统产业的应用和融合而衍生出新型业态





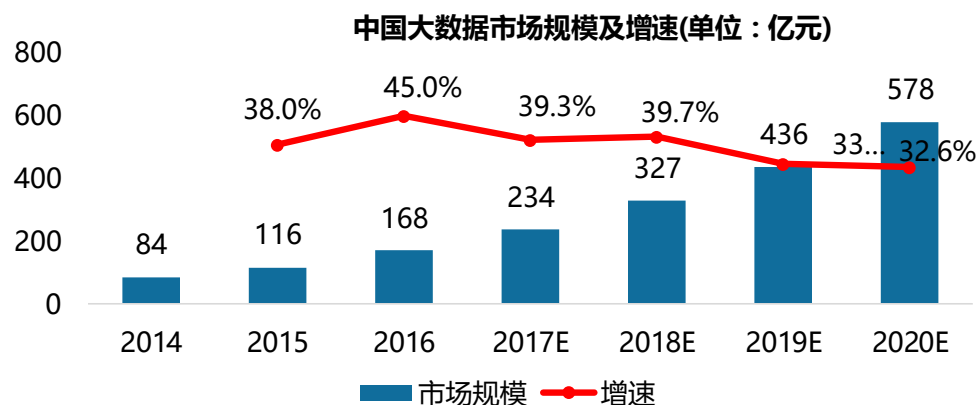
# 全球及我国大数据市场规模（核心产业）



数据来源：Wikibon，2016年3月，单位：亿美元

根据 IDC、Wikibon 等咨询机构分析，2016年，全球的大数据核心产业规模约为300亿美元，2000亿人民币左右。

根据信通院的市场调查，2016年中国大数据市场规模约为168亿人民币。



数据来源：大数据市场调查报告（中国信通院，2017.03）

# 我国大数据企业区域分布



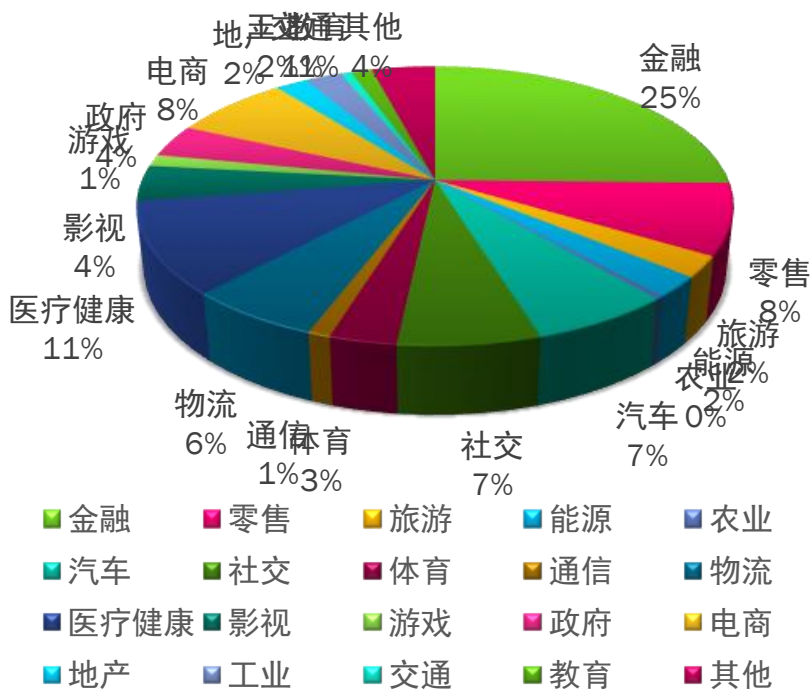
全国大数据企业热力分布图  
(企业数量+叠加注册资本)

数据来源：中国大数据和人工智能产业分析平台（中国信通院，2017.03）

如图所示：北京地区、长三角、珠三角及成渝经济圈是大数据企业热度最高的区域。

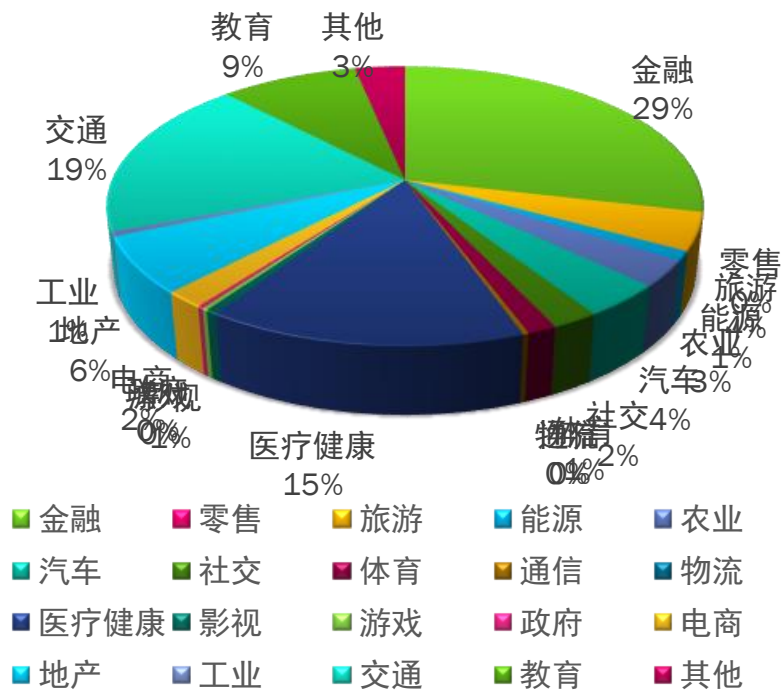
# 大数据应用领域渗透分析

## 行业服务渗透分析



**排名前三的领域：金融、医疗、社交**

## 行业应用渗透分析



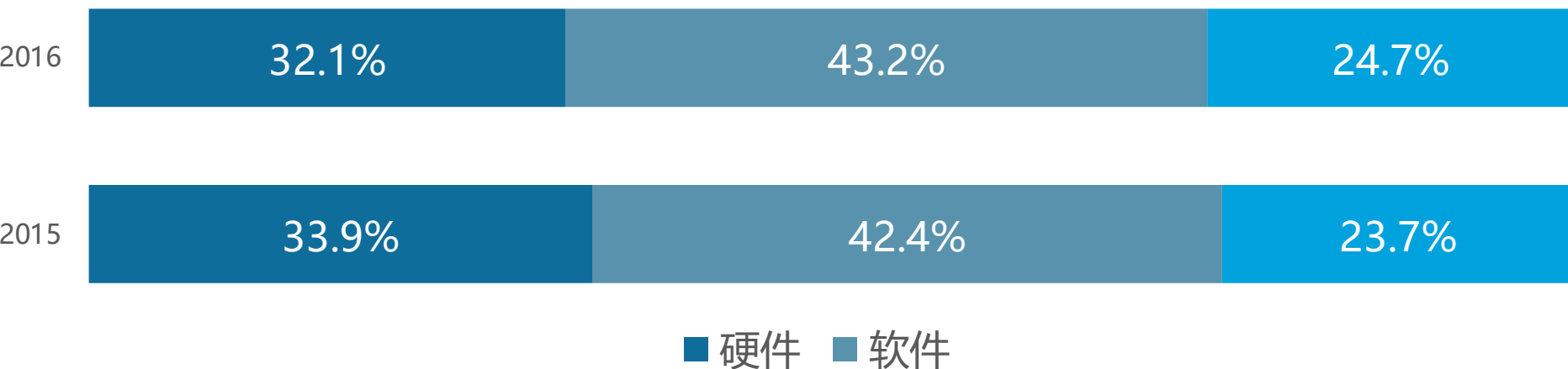
**排名前三的领域：金融、医疗、交通**

数据来源：中国大数据和人工智能产业分析平台（中国信通院，2017.03）

# 大数据细分市场逐年变化

根据中国信息通信研究院的统计，2016年与2015年相比，硬件市场规模为53.9 亿元，占比下降1.8%；软件市场规模为72.6亿元，占比提高0.8%；大数据服务市场规模41.5亿元，占比提高1.0%。这体现出大数据市场细分领域中，软件和服务比重呈上升趋势，硬件比重逐年减少。

中国大数据细分领域构成



## 三、发展大数据的战略选择

- 1、选择先进的大数据技术及产品
- 2、保证使用、流通数据资源时的合规性

# 大数据技术快速演进，底层技术日趋成熟

阶段1：RDBMS一枝独秀  
One size fits all  
SQL/ACID

阶段2：NoSQL百花齐放  
One Size Does Not Fit All  
NoSQL/CAP

阶段3：回归SQL  
One Size Does Not Fit All  
SQL/NewSQL

面向企业应用的商业智能提出，面向数据仓库的数据分析（OLAP）技术兴起

开源Apache Hadoop逐渐兴起，大幅推动互联网大数据应用

Spark、Flink等新一代融合大数据平台，SQL on Hadoop兴起

1995

2000

2005

2010

2015

分布式技术提出  
谷歌提出分布式文件系统、分布式数据库和分布式计算框架，奠定大数据技术基础

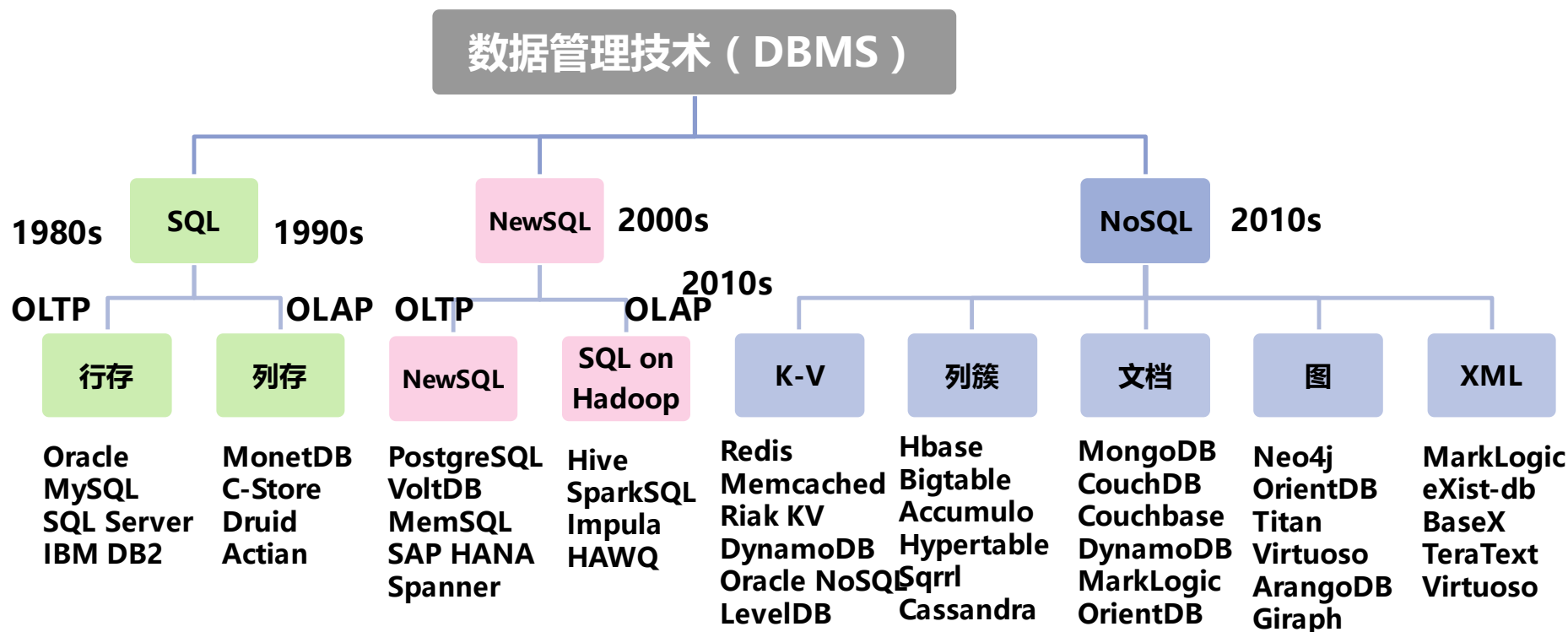
实时计算技术提出  
流计算、图计算、交互式分析、内存计算等技术不断演进

前大数据时代

底层分布式技术日趋成熟，释放上层数据分析技术的创新活力：  
实时流分析、交互分析、机器学习

# 数据平台技术创新活跃，解决方案百花齐放

大数据需求拉动技术创新，衍生出310多种不同的数据管理技术



**对上采用统一接口，底层使用内存技术成为主要趋势：**

1. SQL再次流行，大数据平台向数据管理平台发展，SQL on Hadoop技术成为
2. SSD和Flash广泛使用，颠覆了之前数据库设计的假设，内存数据库发展迅猛，如HAHA，Spark
3. 交易（OLTP）型和分析（OLAP）型相互融合，产生新的融合数据管理平台技术（HTAP）
4. 各种NoSQL技术百花齐放，one size dose not fit all



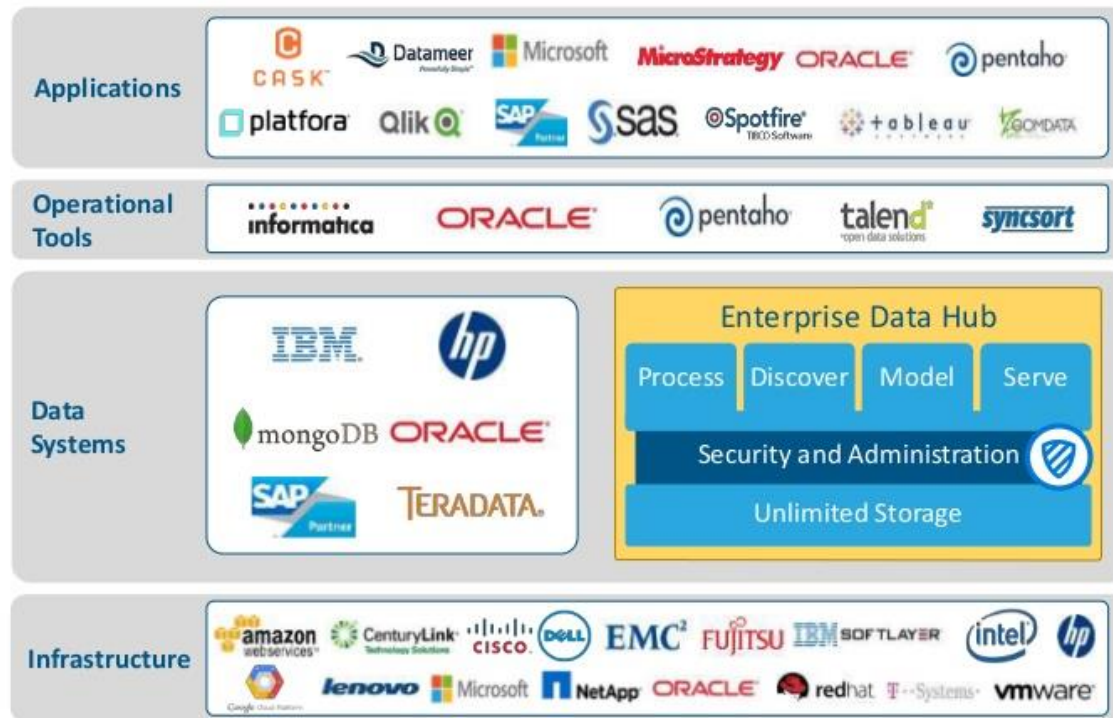
# 大数据平台产品竞争激烈，进入拼生态阶段

未来只有少数平台存活，竞争日益激烈

Hadoop平台产品前四强：Cloudera、Hortonworks、MapR和IBM

如：Cloudera不断扩大兼容性列表，Hortonworks倡导Hadoop的标准化

None of this is possible without the ecosystem



More than 1,600 partners ensure compatibility with existing investments, lower skill barriers, and help maximize value from your data.



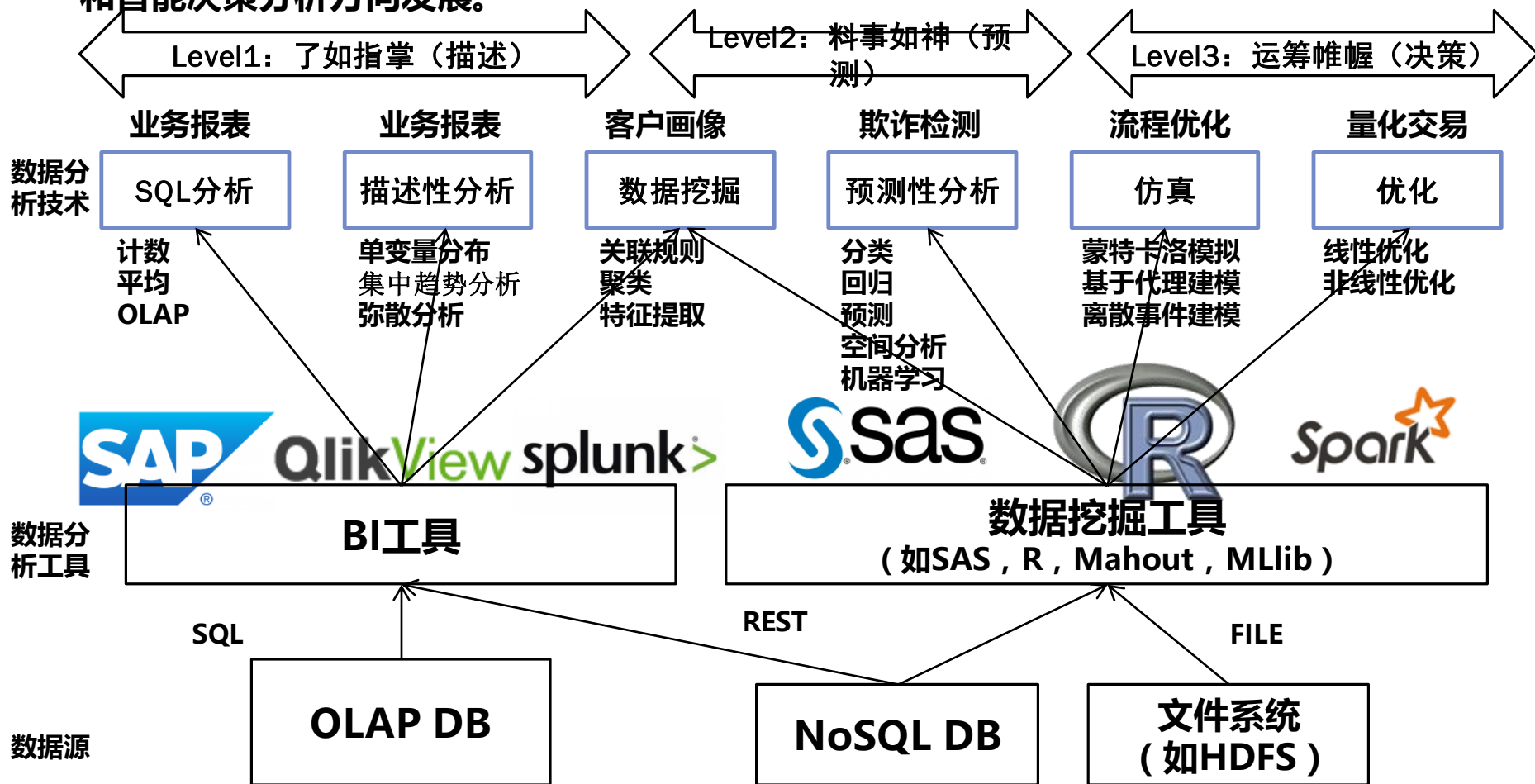
cloudera

© Cloudera, Inc. All rights reserved. 16









# 大数据分析技术：从了如指掌到运筹帷幄

大数据分析技术：正在从基于统计方法的描述性分析，向基于机器学习的智能预测分析和智能决策分析方向发展。



# 大数据技术平台的选型需要公平的衡量标准

## 开源软件→ 商用产品

	可用性：包括系统的稳定性和高容错性能力等
	架构兼容性：与原有系统的兼容性（SQL接口），与开源架构的兼容性
	性能：高吞吐量和性价比，水平线性扩展能力
	安全性：隐私数据的加密，严格的权限管理，完善的数据保护
	易用性：快速的集群部署以及扩展，自动调优，自动化平台升级
	灾备能力：业务连续性和监管合规

# 大数据基础平台产品逐步走向成熟

运维管理	可用性	功能	兼容性	安全	多租户	扩展性
自动化部署	Namenode主节点失效恢复	数据导入	ODBC兼容性	认证	租户管理	集群动态扩展
资源监控	Namenode备节点失效恢复	SQL任务能力	JDBC兼容性	授权	资源管理	集群动态收缩
作业监控	Datanode节点失效恢复	NoSQL数据库	SQL支持度	加密	资源隔离	
集群操作	HMaster节点失效恢复	机器学习	传统数据库同步	审计	权限管理	
故障管理	RegionServer节点失效恢复	流处理能力	跨不同数据库表关联操作			
日志管理	HDFS备份恢复					
配置管理	HBase备份恢复					
权限管理	运维管理节点失效及恢复					
用户管理						
无宕机升级						

大数据产品基础能力认证包括七大项：功能、运维、多租户、可用性、安全、兼容性、扩展性，总共**38**项测试用例

# 大数据基础平台能力评测——量化技术实力

数据中心联盟 《大数据基准测试2.0》 **先后四轮测试**

全球首次在**统一的测试平台**上进行测试

国内**第一个工业级大数据**基准测试活动，20多家国内顶级大数据商业发行版厂商

国内**第一个多个任务负载**的大数据基准测试活动，覆盖性能、容错性、安全与运维

国内外**第一次对大数据公有云服务**进行测试

中国电子科技集团公司第二十八研究所

中兴通讯技术有限公司

成都四方伟业软件股份有限公司

杭州泰一指尚科技有限公司

北京中联润通信息技术有限公司

航天信息股份有限公司

联想（北京）有限公司

深圳前海信息技术有限公司

中国华戎科技集团有限公司

28th鲸云大数据存储分析平台软件[鲸云大数据平台]V2.0

大数据平台DAP V2.0

SDC大数据基础平台V3.0

泰一指尚DMP数据聚合与管理服务平台V1.0

易云汇智大数据平台V1.0

航天信息大数据分析平台V1.0

联想企业级数据分析平台[LEAP系统]V3.2

前信闪电大数据系统 V1.0

华擎大数据平台 V2.0

# 大数据基础平台的技术现状

企业	无宕机升级	运维管理节点HA	SQL支持度	跨数据库关联	多租户资源隔离	多租户权限
中电28所	大版本升级	√	93条语句	√	√	√
中兴	小版本升级	√	99个语句	√	√	√
中联润通				√		
泰一指尚	大版本升级	√	99个语句	√	√	√
航天信息			10个语句		√	√
联想	小版本升级	√	99个语句	√	√	√
前海信息	小版本升级	√	10个语句	√	√	√
四方伟业	小版本升级		10个语句	√	√	√
华戎		√	99个语句	√	√	√

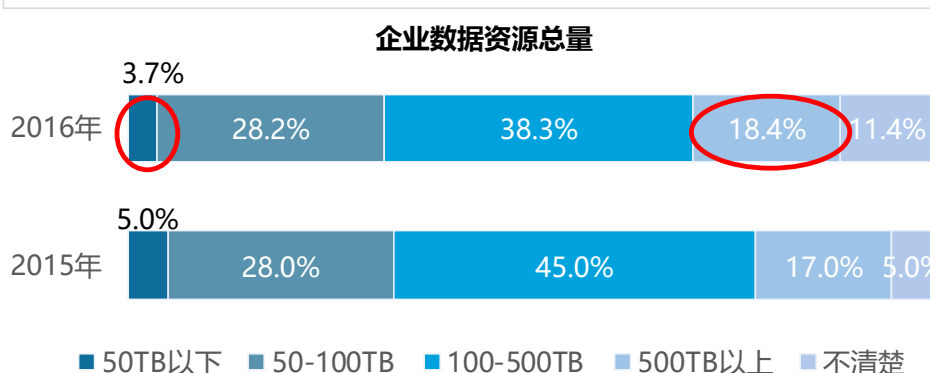
## 三、发展大数据的战略选择

1、选择先进的大数据技术及产品

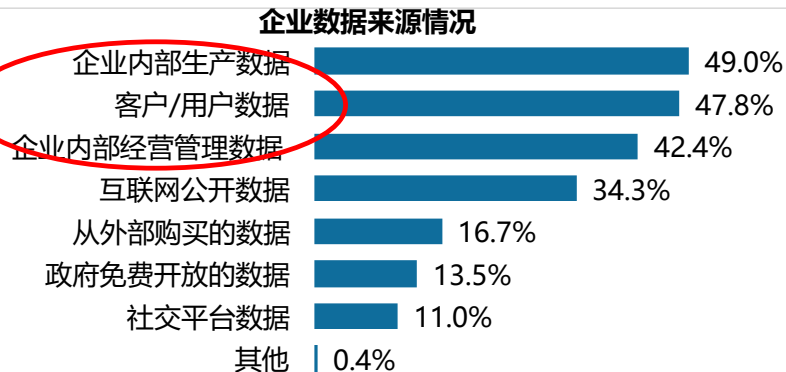
2、保证使用、流通数据资源时的合规性

# 大数据的资源现状

## 企业数据资源规模增长迅速，企业数据资源的主要来自于企业内部

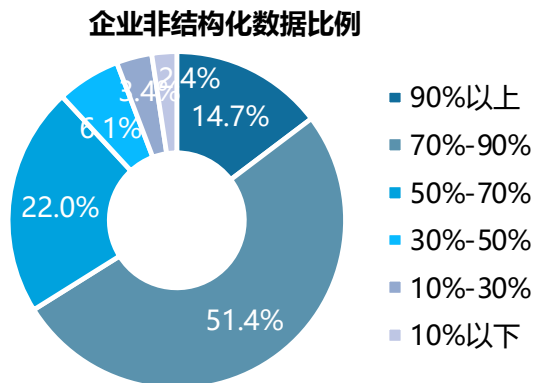


数据来源：中国信息通信研究院 2017,03

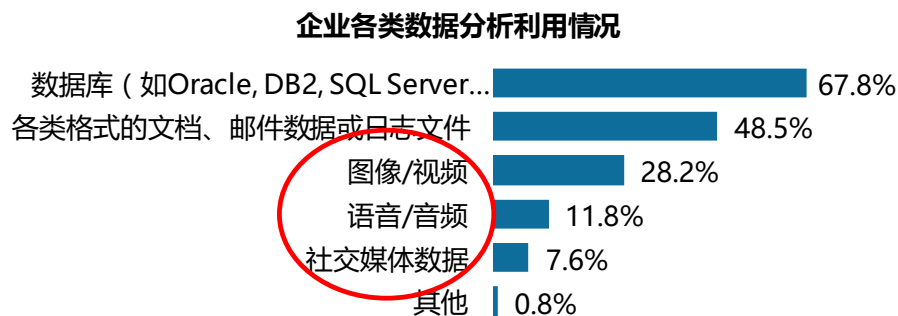


数据来源：中国信息通信研究院 2017,03

## 绝大部分数据资源为非结构化数据，而图像/视频、语音/音频、社交媒体数据等非结构化数据的分析和利用有待提高

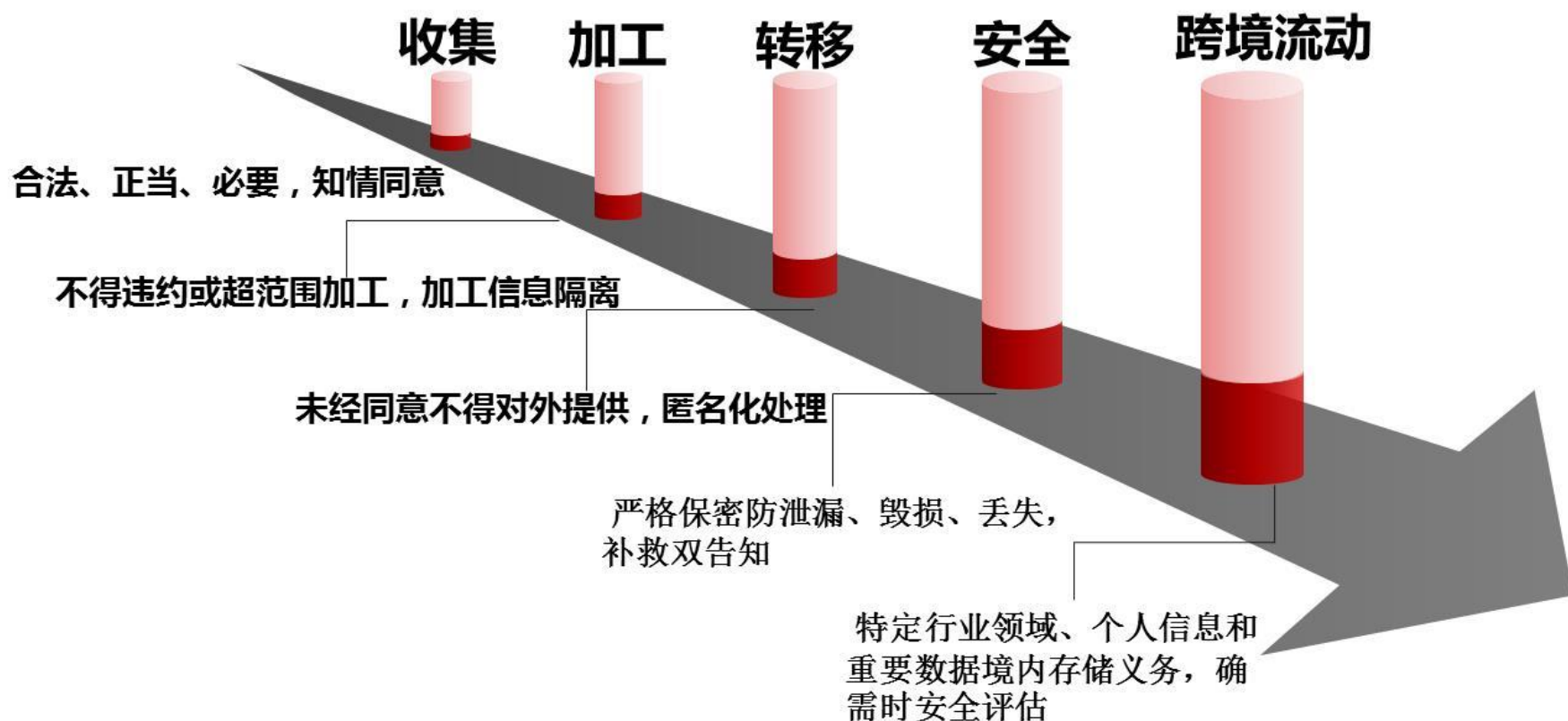


数据来源：中国信息通信研究院 2017,03



数据来源：中国信息通信研究院 2017,03

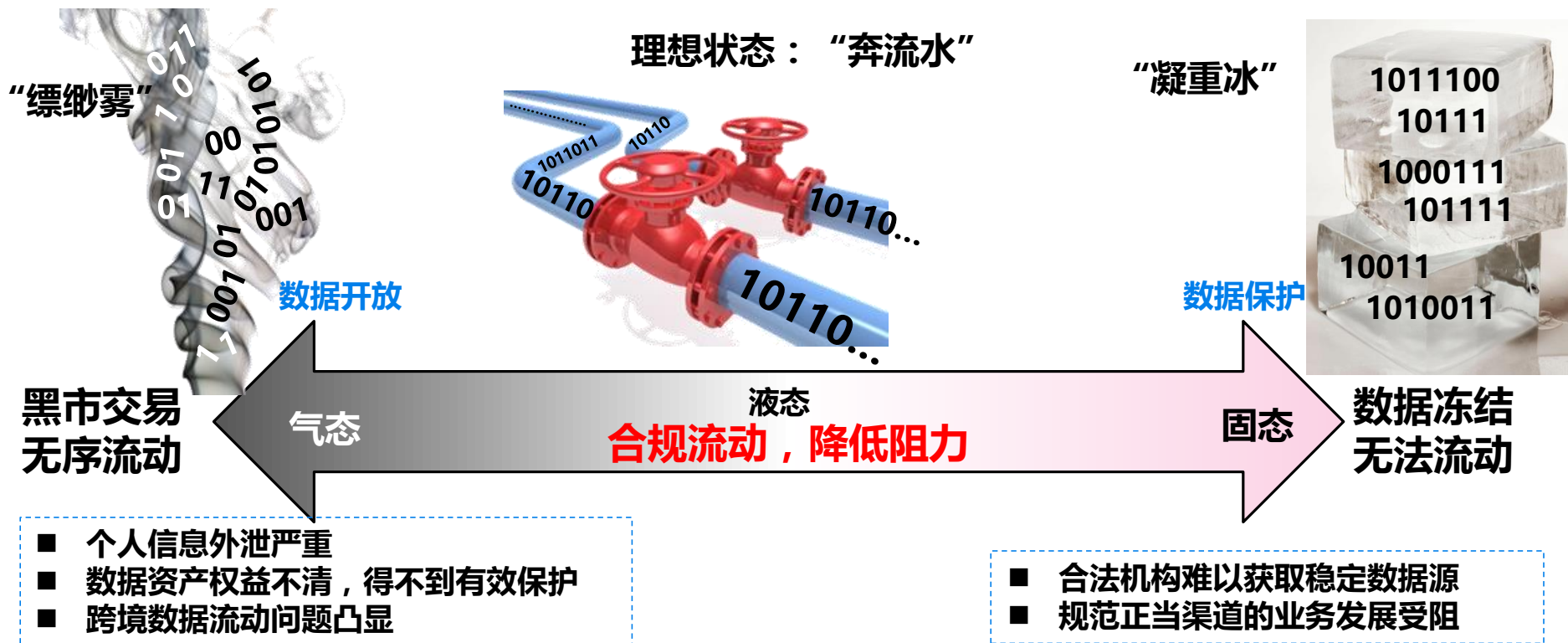
# 大数据存在合规性要求





# 大数据流通交易

## 数据流通交易面临两个极端



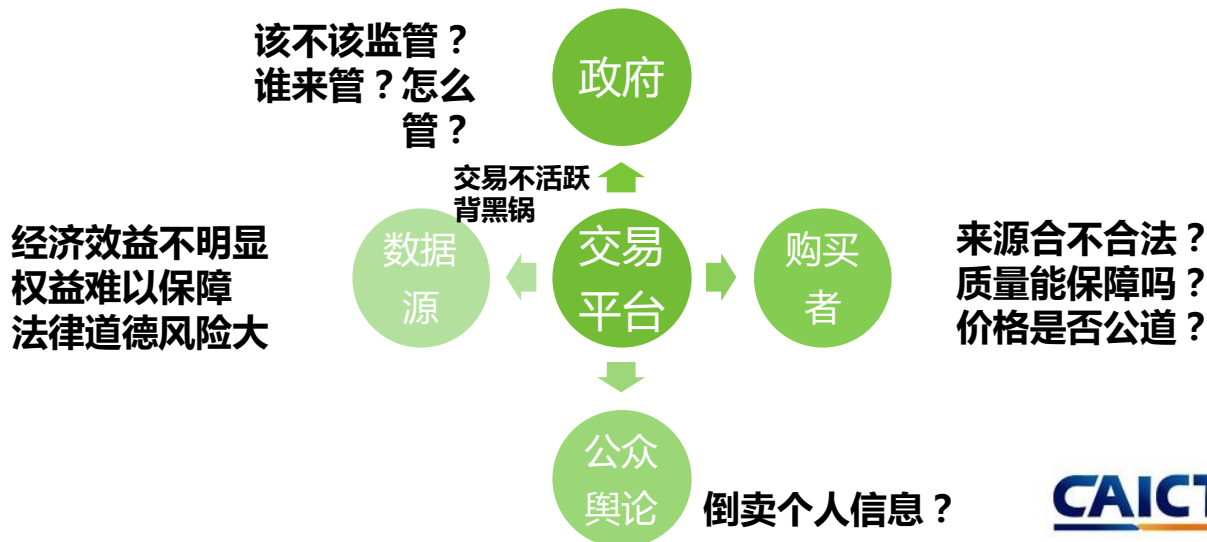
# 数据流通高速发展，各方利益相互交织

## 地方大数据交易平台建设进入井喷期



截至2017年5月，全国已建成18个以上大数据交易场所，到2020年，全国数据交易平台数量预计达到30个

北京	上海	浙江	贵州
重庆	深圳	广州	武汉
贵阳	无锡	湛江	.....



# 数据跨境流动带来安全风险

## 数据泄露风险

境外当地数据保护水平不足，数据泄露

境外当地执法部门执法强制披露数据

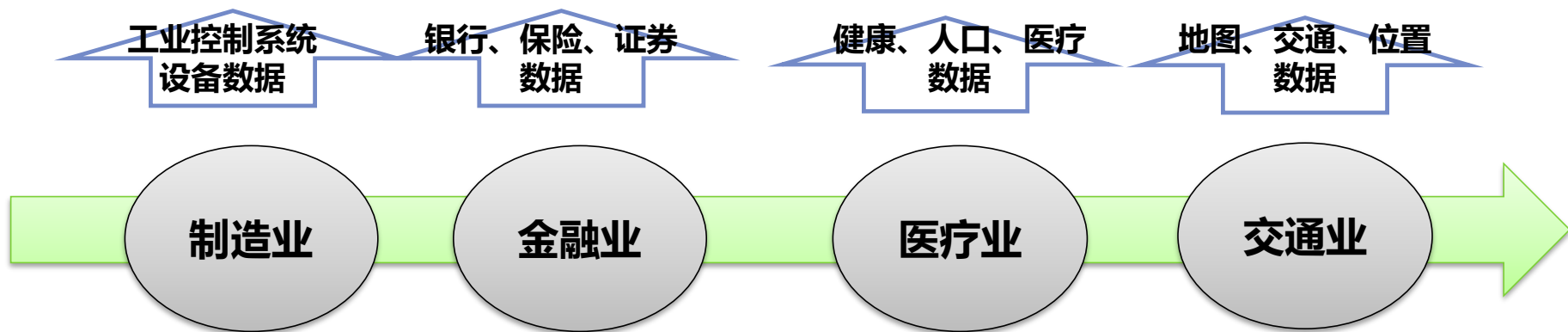
境内侵权行为维权困难

## 丧失数据资源优势

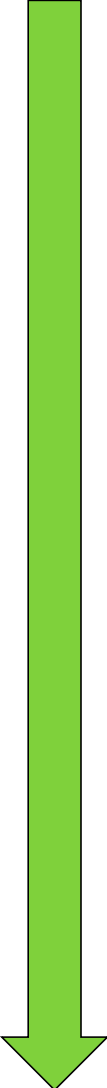
各国信息基础能力差异导致收集、处理、运用数据能力差异

美等强国掌握各国大量关乎经济社会及国家安全的**关键信息**

部分国家形成更强的**信息资源优势**和**战略控制能力**



# 解决思路1：采用完善数据流通安全措施



**技术安全措施：**用于生成特有身份识别信息的算法信息在数据去身份化之后及时销毁，使去身份化不可被撤销。

**业务流程控制：**单位应当制定去身份化流程以及去身份化技术实施方案，并定期审查和更新，以确保流程和方案适当且符合当前的技术水平和行业实践。

**数据分级制度：**建立严密的内控流程，根据数据的敏感层度进行数据敏感分级，针对不同分级，在使用数据前，需要经过相应敏感级别对应所需的审批流程。

**完善管理制度：**制定对于身份关联信息或去身份化信息的管理制度，规范员工行为，并设置专门人员定期对于业务流程进行审核，尽量将数据的使用控制在具有安全防范措施的平台内部，避免敏感数据被调出后存在可能泄露的风险。

# 解决思路2：推进大数据法律制度建设

- 在大数据和“互联网+”背景下，结合我国数据产业发展和行业监管实践，大数据管理立法应该明确“安全与发展并重、促进与规范并举”的战略方向和遵循“积极利用、科学发展、依法管理、确保安全”的方针。

- 释放数据经济活力
- 数据商业利用

立法原则

- 保护数据主权与安全
- 保护个人隐私

## 制定整体性、基础性制度

- 确保数据安全和个人隐私保护，强化数据收集、使用规则；
- 加强跨境数据流动管理，建立风险评估和安全审查机制；
- 明晰数据产权，抓紧确立数据交易规则；
- 推动出台政府数据开放法规政策

## 就重点问题制定针对性制度

- 高价值、高风险的重点数据：用户数据、用户个人信息。
- 涉及数据价值变动和法律关系变化的重点行为：数据开放、数据共享、数据交易、数据承接等。
- 与权益保护相关的制度：泄露通知等。

# 解决思路3：构建行业自律

由中央网信办、国家发改委、工业和信息化部指导，中国信息通信研究院发起，数据流通相关核心企业及大型互联网企业参与制定的“数据流通行业自律公约”已发布。

## 《数据流通行业自律公约》

公约2.0版  
架构

总则

数据  
权益

数据  
流通

数据  
应用

附则

公约1.0版  
内容

公约宗旨

1.遵守法律

2.合规采集

4.数据保护

3.流向管控

5.权益协调

6.服务提升

7.社会监督

新增内容

数据流通  
的定义

企业对数据  
的权益

强化流通环  
节风险评估

企业增强  
社会责任

公约的持续  
完善更新

完善内容

个人、企业  
的利益协调

第三方平台  
责任义务

推动技术标准  
与认证、审计

成立日常监  
督管理机构

# 解决思路4：标准化为数据流通提供保障

目前，数据中心联盟正在制订一系列数据流通相关标准，并委托中国信息通信研究院进行标准符合性测试。

安全管理 机制	数据使用 规范	身份识别	信息验证	数据转移 方式	授权方式 及流程	接口	审计
------------	------------	------	------	------------	-------------	----	----

征信类产品及服务	第一批测试已完成
----------	----------

金融风控类产品及服务	第一批测试已完成
------------	----------

数据流通中心数据处理及业务管理	第一批测试已完成
-----------------	----------

精准营销类产品及服务	标准制订中
------------	-------

位置信息类产品及服务	标准待制订
------------	-------



# 数据流通平台合规性现状

## 首批通过**征信服务类数据流通标准**符合性测试的企业



中国电信股份有限公司云计算分公司



中国联通信息化事业部数据中心

## 首批通过**金融风控类数据流通标准**符合性测试的企业



四川享宇金信金融服务外包公司



北京通付盾数据科技有限公司

## 首批通过**数据流通中心数据处理及业务管理标准**测试的企业



上海数据交易中心



京东万象



安徽大数据交易中心



# 《网络安全法》立法进程

- 我国近年来加快了网络安全方面的立法步伐。2015年7月1日，十二届全国人大常委会第十五次会议通过《国家安全法》，明确要求“维护国家网络空间主权、安全和发展利益”，“实现网络和信息核心技术、关键基础设施和重要领域信息系统及数据的安全可控”，网络安全成为国家安全的重要组成部分。
- 《网络安全法》三读如期通过。在《国家安全法》的大框架之下，为解决我国网络安全领域存在的突出问题，以制度建设提高国家网络安全保障能力，掌握网络空间治理和规则制定方面的主动权，切实维护国家网络空间主权、安全和发展利益，顺应国内外日益复杂的网络安全态势，迫切需要在国家层面制定一部统一的网络安全基本法。在此背景下，2015年7月6日，全国人大网公布《网络安全法》（草案），向社会公开征求意见。2016年7月5日，全国人大第二次公布《网络安全法》（草案），向社会公开征求意见。
- 2016年11月7日全国人大常委会正式出台《网络安全法》，自2017年6月1日起施行。

# 《网络安全法》主要内容

序号	主要制度内容
(一)	网络空间主权和网络安全战略 (第1、4条)
(二)	网络安全监管体制 (第8条)
(三)	网络安全等级保护制度 (第21条)
(四)	网络关键设备和网络安全专用产品安全认证检测制度 (第23条)
(五)	网络实名制的网络身份管理制度和网络可信身份战略 (第24条)
(六)	合法侦听协助制度 (第28条)
(七)	关键信息基础设施运行安全保护制度 (第31-34条、第38条、第39条)
(八)	关键信息基础设施网络产品服务安全审查制度 (第35条)
(九)	关键信息基础设施重要数据境内留存制度 (第37条)
(十)	个人信息保护制度 (第40-44条)
(十一)	网络信息管理制度 (第46-49条)
(十二)	重大突发社会安全事件的网络通信临时限制制度 (第58条)

# 逐步完善网络安全立法

## 《网络安全法》是网络安全立法的新起点而非终点。

### ● 为建设网络强国提供制度保障：

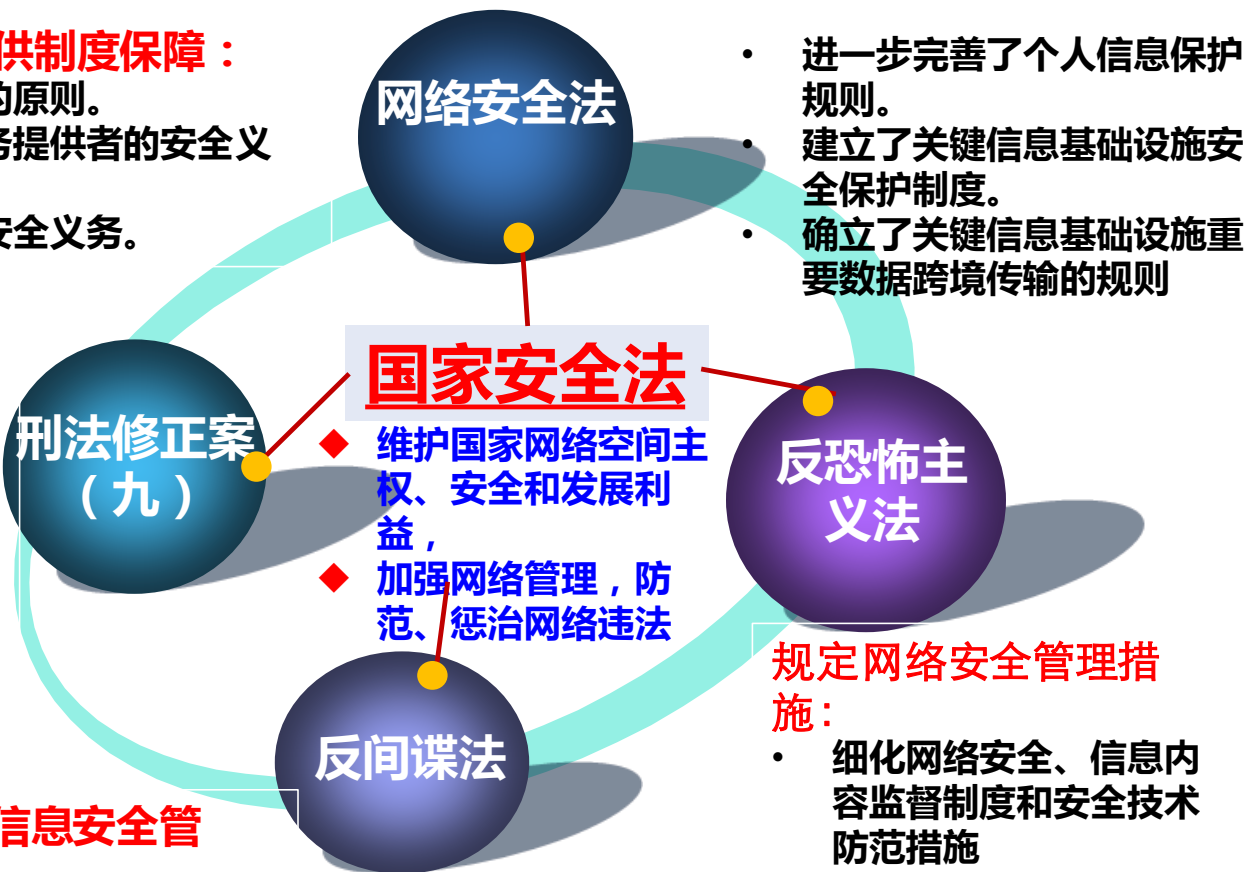
- 明确了网络空间主权的原则。
- 明确了网络产品和服务提供者的安全义务。
- 明确了网络运营者的安全义务。

### ● 护航信息网络安全：

- 加强对公民个人信息的保护
- 明确ISPS履行网络安全管理的义务
- 完善网络犯罪规定
- 增加编造、传播虚假信息犯罪的规定

### ● 明确间谍行为强化信息安全管理：

- 第十三条：国家安全机关因反间谍工作需要，可以依照规定查验有关组织和个人的电子通信工具、器材等设备、设施。



# 《网络安全法》确立了个人信息基本保护制度

规定了相关主体的个人信息保护义务，为大数据“划底线”

## ◆ 对于网络运营者等企业主体而言，要求：

- ✓ 在收集使用个人信息的时候遵守《网络安全法》规定的基本原则（第41条）；
- ✓ 未经被收集者同意，不得向他人提供个人信息（第42条）；
- ✓ 个人信息的更正和删除义务（第43条）
- ✓ 应当建立网络信息安全投诉、举报制度，及时受理并处理有关网络信息安全的投诉和举报（第49条第1款）；
- ✓ 并积极配合网信部门和有关部门依法实施的监督检查（第49条第2款）。

## ◆ 网络安全监督管理部门及其工作人员，要求：

- ✓ 必须对在履行职责中知悉的个人信息、隐私和商业秘密严格保密，不得泄露、出售或者非法向他人提供。此外，任何个人和组织不得窃取或者以其他非法方式获取个人信息，不得非法出售或者非法向他人提供个人信息（第50条）。

# 《网络安全法》为大数据流通留下发展空间

《网安法》严格保护个人数据，但也通过“匿名化”等规定为大数据的流通提供了可能

## ◆ 在数据流通方面，

第四十二条虽然要求“未经被收集者同意，不得向他人提供个人信息”，但本款也规定了例外情形，即“经过处理无法识别特定个人且不能复原的除外”。个人信息数据匿名化处理是维护数据安全的有效手段，也是兼顾隐私保护和数据利用的有效路径。

## ◆ 在数据交易方面，

《人大决定》明确规定“不得出售”公民个人信息，但《网络安全法》第四十四条规定“任何个人和组织不得非法出售或者非法向他人提供个人信息”，也就是说，在一定合法情形下是可以出售和提供公民个人信息的，这无疑给符合规定的个人信息交易开了绿灯，为我国大数据产业发展提供了空间。

# 两高解释带来企业数据合规的新挑战



在《网络安全法》2017年6月1日正式生效前，最高人民法院与最高人民检察院于2017年5月9日发布《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》（简称“《两高解释》”）及相关典型案例并将与《网络安全法》同时生效。

《两高解释》第一次对“侵犯公民个人信息罪”的犯罪要件做出了较为明确的规定，对于明确刑事责任的范畴和判断具有重要意义。

- 明确“违反国家有关规定”范围
- 明确未经同意提供公民个人信息构成犯罪
- “以其他方式非法获取公民个人信息”的概念扩宽

《两高解释》和《网络安全法》的生效，将大大提高企业在数据保护、网络安全保护等方面的合规义务

# 大数据企业《网络安全法》合规性指引（1）

## 1. 对个人信息数据进行脱敏处理可实现“合规”

在数据流通方面，第四十二条要求“未经被收集者同意，不得向他人提供个人信息”，但“经过处理无法识别特定个人且不能复原的除外”。**网络安全法把“个人隐私”和“数据信息”加以了区别式定义**，即“经过处理无法识别特定个人且不能复原的”的才可流通。大数据企业应在此条款的指导下，合法开展大数据应用。



# 大数据企业《网络安全法》合规性指引（2）

## 2. 明确信息收集和使用的目的、方式、范围可实现“合规”

《网络安全法》第四十一条规定“网络运营者收集、使用个人信息，应当遵循合法、正当、必要的原则，公开收集、使用规则，明示收集、使用信息的目的、方式和范围，并经被收集者同意”。因此，大数据企业在进行数据收集和使用时，需做到明示收集、使用信息的目的、方式和范围，并经被收集者同意。

# 大数据企业《网络安全法》合规性指引（4）

## 3.关键信息境内存储可实现“合规”

第三十七条要求“关键信息基础设施的运营者在中华人民共和国境内运营中**收集和产生的个人信息和重要数据**应当在境内存储。因业务需要，确需向境外提供的，应当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；法律、行政法规另有规定的，依照其规定。”因此，大数据企业应更加关注自身数据的存储安全。一般情况下应确保收集和产生的数据在境内存储。

# 大数据企业《网络安全法》合规性指引（3）

## 4.重要系统和数据库容灾备份可实现“合规”

第三十三、三十四条要求“关键信息基础设施的运营者，应当确保其具有支持业务稳定、持续运行的性能”，并通过设置专门安全管理机构和安全管理负责人；定期对从业人员进行网络安全教育、技术培训和技能考核；**对重要系统和数据库进行容灾备份**；制定网络安全事件应急预案等方式履行安全保护义务。因此，对于关键信息基础设施的运营者来说，对重要系统和数据库进行容灾备份是必要且重要的。



厚德實學 興業致遠

政府高端专业智库  
行业创新发展平台

韩涵  
13811249349  
[hanhan@ritt.cn](mailto:hanhan@ritt.cn) ; [hanhan@caict.ac.cn](mailto:hanhan@caict.ac.cn)



扫一扫上面的二维码图案，加我微信